

UNIVERSITA' DEGLI STUDI DI NAPOLI

"FEDERICO II"



TESI DI DOTTORATO

STATISTICA COMPUTAZIONALE

XXI° CICLO

**Il ruolo dei modelli ad equazioni strutturali basato
sull'entropia per l'analisi della Customer Satisfaction:**

**Un Confronto tra il Partial Least Squares (PLS) e
Generalized Maximum Entropy (GME)**

Applicazione nella Grande Distribuzione

COORDINATORE

Ch.mo Prof. Carlo Natale LAURO

CANDIDATO

Dott. Andrea CALIGIURI

SUPERVISORE

Dottor Enrico CIAVOLINO

INDICE

CAPITOLO 1. La Customer Satisfaction Nella Prassi Aziendale Con Particolare riferimento alla Grande Distribuzione

Premessa

1.1 Il Ruolo Della Customer Satisfaction nella Vision 2000

1.2 La Revisione Dello Standard e Le Principali Novità: ISO 9001:2008

1.3 Customer Satisfaction e Total Quality Management: Quale interazione?

1.3.1 Focalizzazione sul cliente

1.3.2 Sviluppo e coinvolgimento del personale

1.3.3 Misurazione della qualità

1.3.4 Miglioramento continuo

1.4 La Customer Satisfaction come strumento gestionale

1.4.1 L'analisi della Customer Satisfaction nell'ambito del processo di "Ascolto del cliente"

1.4.2 I reclami dei clienti

1.4.3 Analisi dei clienti perduti

1.4.3. Le indagini sui clienti misteriosi

1.5 Il processo di formazione della Customer Satisfaction

1.6 Le tre principali caratteristiche della soddisfazione: soggettività, relatività ed evoluzione

1.7 Analisi delle Fidelity card

CAPITOLO 2. Modelli Per La Misurazione Della Customer Satisfaction

2.1 Misurare la Customer Satisfaction

2.2 Metodologie utilizzate per la valutazione della soddisfazione nella Grande distribuzione

2.2.1 Nei modelli diretti formativi o compositivi

2.2.2 Modelli Diretti Esplicativi o Decompositivi

2.2.3 Modelli Di Reti Bayesiane

2.2.4 Modelli Strutturali

CAPITOLO 3. Equazioni strutturali: Formulazioni del modello

Premessa

3.1 Tipi di relazioni causali tra le variabili.

3.2.1 Relazione diretta.

3.2.2 Relazione reciproca.

3.2.3 Relazione spuria.

3.2.4 Relazione indiretta.

3.2.5 Relazione condizionata (interazione)

3.3 «Lisrel» ed i modelli di equazioni strutturali.

3.4 La logica e le fasi di Lisrel.

3.5 La formulazione del modello secondo la notazione Lisrel.

3.6 La rappresentazione grafica del modello.

3.7 Il modello strutturale.

3.8 Il modello di misurazione per le variabili esogene.

3.9 Il Modello Di Misurazione Per Le Variabili Esogene.

3.10 Le restrizioni del modello

CAPITOLO 4. Stima dei parametri e verifica empirica del modello ad equazioni strutturali.

4.1 La logica del procedimento di stima dei parametri del modello.

4.2 La covarianza fra le variabili esogene X.

4.3 La covarianza fra le variabili endogene Y.

4.4 La covarianza fra le variabili endogene Y e le esogene X.

4.5 La stima dei parametri strutturali

4.6 L'adattamento del modello ai dati.

4.7 Le covarianze espresse in funzione dei parametri.

4.8 Regole di scomposizione delle covarianze e delle varianze.

4.9 Le misure di adattamento complessivo del modello.

4.10 Il miglioramento del modello.

4.10.1 Esclusione di parametri (valori - t).

4.10.2 Inclusione di nuovi parametri (indici di modifica).

4.11 Stima PLS

CAPITOLO 5. Entropia

5.1 L'Entropia di Shannon come misura della Casualità

5.2 La Massima Entropia Generalizzata

5.3 I modelli ad equazioni strutturali di Massima Entropia Generalizzata

5.4 L'Algoritmo GME per SEM

5.5 Simulazione

CAPITOLO 6. Caso studio

6.1 Lo Studio della CS nel Punto Vendita

6.2 Modello di valutazione della CS

6.3 Confronto PLS - MSE

6.4 Matrice Degli Interventi

6.5 Supporto Alle Decisioni

6.6 Criteri di Simulazione

6.7 Confronto MSE variabili latenti: CASO GENERALE

6.7.1 Confronto MSE variabili manifeste: CASO GENERALE

6.8 Confronto MSE variabili latenti: MISSING VALUE

6.8.1 Confronto MSE variabili manifeste: MISSING VALUE

6.9 Confronto MSE variabili latenti: OUTLIERS

6.9.1 Confronto MSE variabili manifeste: OUTLIERS

6.10 Confronto MSE variabili latenti: MULTICOLLINEARITA' 0,3

6.10.1 Confronto MSE variabili latenti: MULTICOLLINEARITA' 0,3

6.11 Confronto MSE variabili latenti: MULTICOLLINEARITA' 0,9

6.11.1 Confronto MSE variabili latenti: MULTICOLLINEARITA' 0,9

6.12 Analisi Dei Risultati

APPENDICE

BIBLIOGRAFIA

Ringraziamenti

Al termine di questo percorso di studi, doverosi sono i ringraziamenti alle persone che mi hanno seguito durante questi anni.

In primo luogo desidero ringraziare il Prof. C. Lauro che con il suo aiuto, i suoi preziosi consigli mi ha spronato ad andare avanti anche quando tutto sembrava irrisolvibile. Grazie di cuore per la sua disponibilità e per l'opportunità datami.

A colui che ha guidato, limato, arricchito il mio lavoro con la sua conoscenza, i suoi suggerimenti, il dott. E. Ciavolino un sentito grazie.

Inoltre desidero ringraziare chi mi ha iniziato a questo percorso di studi, il Prof D'Ambra

Amici e colleghi con i quali ho condiviso giorni di studio e momenti di confronto, siete sempre nei miei ricordi nonostante le scelte diverse.

CAPITOLO PRIMO

La Customer Satisfaction Nella Prassi Aziendale Con Particolare riferimento alla Grande Distribuzione

Sommario: Premessa - **1.1** Il Ruolo Della Customer Satisfaction nella Vision 2000 - **1.2** La Revisione Dello Standard e Le Principali Novità: ISO 9001:2008 - **1.3** Customer Satisfaction e Total Quality Management: Quale interazione? - **1.3.1** Focalizzazione sul cliente - **1.3.2** Sviluppo e coinvolgimento del personale - **1.3.3** Misurazione della qualità - **1.3.4** Miglioramento continuo - **1.4** La Customer Satisfaction come strumento gestionale - **1.4.1** L'analisi della Customer Satisfaction nell'ambito del processo di "Ascolto del cliente" - **1.4.2** I reclami dei clienti - **1.4.3** Analisi dei clienti perduti - **1.4.3.** Le indagini sui clienti misteriosi - **1.5** Il processo di formazione della Customer Satisfaction - **1.6** Le tre principali caratteristiche della soddisfazione: soggettività, relatività ed evoluzione - **1.7** Analisi delle Fidelity card

Premessa

L'adempimento degli obblighi normativi legati alla recente evoluzione della ISO 9000 fino alla ISO 9001:2008 e la volontà di accrescere i profitti futuri rappresentano le due principali motivazioni che generalmente inducono un'azienda a misurare la soddisfazione dei propri clienti.

L'ente di normazione francese (AFNOR) definisce la Customer Satisfaction (CS) come "l'opinione di un cliente risultante dallo scarto tra la sua percezione di un prodotto o servizio consumato e le sue aspettative" (ISO/DIS 9000, marzo 1999).

Partendo da questa definizione, scopo del presente capitolo sarà quello di analizzare gli aspetti salienti che consentono di calare la CS nella prassi aziendale, ovvero il legame tra CS e Certificazione e Qualità Totale (QT), la definizione della CS come strumento strategico della gestione aziendale, le caratteristiche della CS ed infine le peculiarità della CS nella grande distribuzione.

1.1 Il Ruolo Della Customer Satisfaction nella Vision 2000

L'evoluzione delle ISO 9000 è fortemente collegata all'evoluzione dei bisogni delle aziende e della disciplina della qualità e delle sue applicazioni. Le norme infatti non rappresentano la perfezione o lo stato dell'arte della disciplina: le norme sono, per definizione, l'opportuno compromesso deciso dalla maggioranza dei paesi partecipanti alla loro costruzione.

Uno dei primi obiettivi che si è posto il comitato tecnico dell'organizzazione internazionale di normazione ISO/TC 176, preposto all'aggiornamento della normativa sulla certificazione, è stato quello di ridimensionare la estesa proliferazione di norme e linee guida ISO. Con tale progetto di revisione, conosciuto sotto il nome di progetto VISION 2000, si è passati dalle oltre 20 pubblicazioni a sole 4 norme di base:

- ISO 9000:2000 – Fondamenti e terminologia;
- ISO 9001:2000 – Sistema di Gestione della Qualità–Requisiti;
- ISO 9004:2000 – Sistema di gestione della Qualità – Linee guida per il miglioramento delle prestazioni;
- ISO 19011 – Verifiche Ispettive

La ISO 9001 e la ISO 9004 sono le norme che definiscono, seppur sotto ottiche differenti, i Sistemi di Gestione per la Qualità (SGQ). La prima

ne definisce i requisiti e rappresenta la norma di riferimento per rapporti contrattuali e per le certificazioni dei SGQ; la seconda sposta l'ottica del SGQ verso il miglioramento delle prestazioni da parte delle organizzazioni che le adottano, fornendo linee guida e strumenti operativi orientati alla efficienza oltre che alla efficacia e tenendo conto non solo dei clienti ma anche di tutte le parti coinvolte nella prassi aziendale (Stakeholder). Tale ampliamento di prospettiva favorisce l'introduzione di metodologie gestionali ancora più avanzate, quali il Total Quality Management (TQM) o quelle sottese dai Premi Qualità. Una scelta che ha orientato tutta la stesura delle nuove norme è stata quella di concepire le due norme sui SGQ, ISO 9001 e ISO 9004, come una coppia coerente. Tale coerenza si esplicita in due direzioni: verticalmente, in quanto le norme hanno la stessa struttura in termini di paragrafazione e orizzontalmente perché i paragrafi omologhi delle due norme sono più compatibili e congruenti rispetto a quanto lo fossero prima. L'importanza di tale scelta sta nel fatto che l'utilizzazione congiunta delle due norme facilita l'adozione del SGQ, in quanto la 9004 ha la funzione di fare comprendere le motivazioni, di fornire utili esplicitazioni ed esemplificazioni integrando di fatto, con le sue indicazioni, i requisiti della 9001. Nella stessa ottica esplicativa ed esemplificativa il comitato di revisione ha ritenuto opportuno accompagnare la nuova versione con la stesura di alcuni principi definiti "*Principi di buona gestione*". Tali 8 principi, che riflettono le più moderne impostazioni delle tecniche organizzative, sono:

- Organizzazione orientata al Cliente;
- Leadership;
- Coinvolgimento del personale;
- Approccio basato sui processi;
- Approccio sistemico della gestione;
- Miglioramento continuo;
- Decisioni basate su dati di fatto;

- Rapporto di reciproco beneficio con i fornitori;

La vera novità dell'architettura delle ISO 9001/2000 sono, rispetto alle precedenti edizioni, l'orientamento al cliente e l'approccio basato sui processi. In essa i tradizionali 20 punti, ossatura della ISO 9000:ed. 94 (da 4.1 a 4.20), non sono più considerati a sé stanti, ma inseriti nella gestione dell'intero processo aziendale, appare, pertanto, evidente il collegamento con il principio espresso da Deming basato sulle quattro fasi fondamentali della gestione aziendale: ***Plan-Do-Check-Act*** (PDCA). Nella nuova norma si individuano così quattro macro-attività in cui identificare tutti i diversi processi aziendali, al fine di poterne migliorare l'organizzazione e quindi accrescerne le prestazioni: Responsabilità della Direzione; Gestione delle Risorse; Gestione dei processi; Misura, analisi e miglioramento. Queste quattro macro-attività definiscono rispettivamente i paragrafi 5, 6, 7 e 8 della norma (fig. 1.1).

0 – Introduzione	
1 – Campo di Applicazione	
2 – Riferimenti Normativi	
3 – Termini e Definizioni	
4 – Sistema di Gestione per la Qualità	
4.1 Requisiti generali	
4.2 Requisiti generali per la documentazione	
5 – Responsabilità del vertice dell'organizzazione	
5.1 Impegni del vertice	
5.2 L'attenzione verso il cliente	
5.3 Politica della Qualità	
5.4 Pianificazione	
5.4.1 Obiettivi della Qualità	
5.4.2 Pianificazione della Qualità	
5.5 Amminisrazione	
5.5.1 Generalità	
5.5.2 Responsabilità ed autorità	
5.5.3 Rappresentante del vertice	
5.5.4 Comunicazioni Interne	
5.5.5 Manuale della Qualità	
5.5.6 Gestione dei documenti	
5.5.7 Gestione delle registrazioni della Qualità	
5.6 Riesame da parte del vertice	
6 – Gestione delle Risorse	
6.1 Messa a disposizione delle Risorse	
6.2 Risorse Umane	
6.3 Risorse di Supporto	
6.4 Ambiente di lavoro	
7 – Realizzazione del prodotto e/o servizio	
7.1 Pianificazione dei processi realizzativi	
7.2 Processi relativi al clienti	
7.3 Progettazione e sviluppo	
7.4 Approvvigionamento	
7.5 Produzione ed erogazione servizio	
7.6 Gestione dei dispositivi di misura e monitoraggio	
8 – Misure, analisi e miglioramento	
8.1 Pianificazione	
8.2 Misure e monitoraggio	
8.3 Gestione delle non conformità	
8.4 Analisi dei dati	
8.5 Miglioramento	

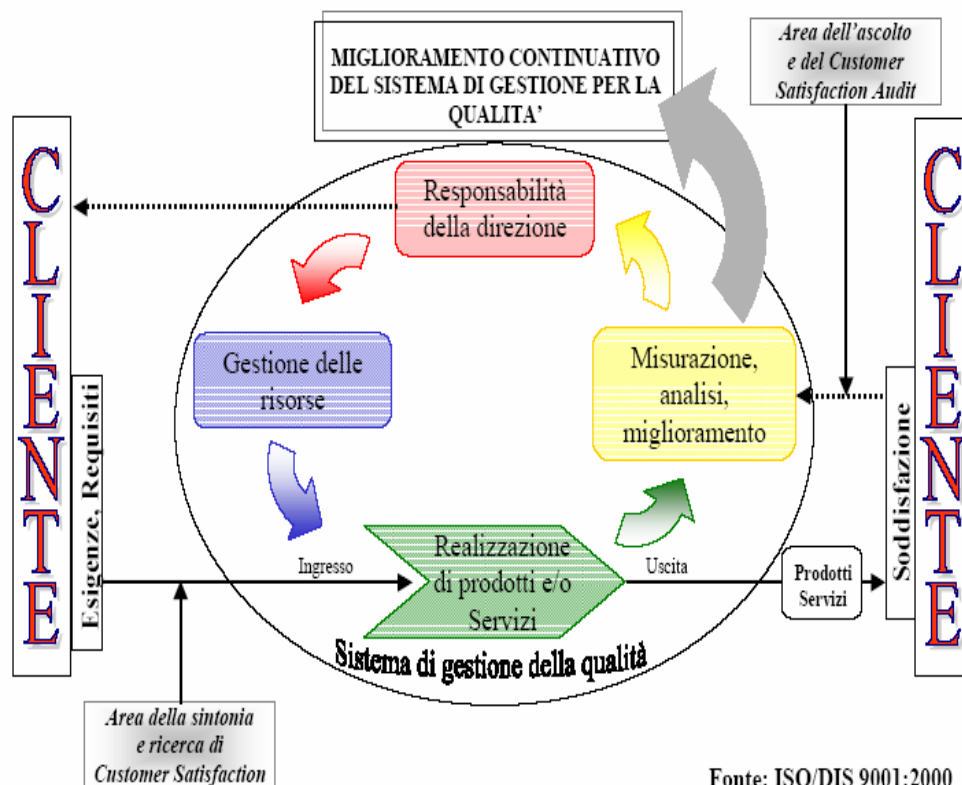
Figura 1.1: La Struttura della ISO 9001:2000

Per quanto concerne il legame tra le quattro attività, rappresentative dei processi aziendali, e il ciclo PDCA è abbastanza evidente trarre quanto segue:

- **PLAN** = Responsabilità della Direzione - Gestione delle risorse;
- **DO** = Gestione dei processi;
- **CHECK** = Misure, Analisi, Miglioramento;
- **ACT** = Misure, Analisi, Miglioramento.

Pertanto, i processi sono ora visti come elementi di un ciclo chiuso, tale propensione risulta essere perfettamente in linea con le più moderne impostazioni organizzative che tendono a ridimensionare il ruolo delle funzioni aziendali a favore dei processi stessi; questa valorizzazione dei processi parte dalla considerazione che sono i processi ad aggiungere valore all'organizzazione. Lo stesso sistema azienda è rappresentato come un macro-processo esplicitabile attraverso 2 cicli (fig. 1.2). Un primo ciclo, detto evolutivo ed interno all'organizzazione, mira sia ad assicurare la conformità alle specifiche del prodotto/servizio, sia a sviluppare la logica del miglioramento continuo, sia alla soddisfazione del cliente. Tale ciclo, partendo dalla responsabilità del management ad individuare e mettere a disposizione adeguate risorse finanziarie ed umane, si sviluppa nella gestione dei processi realizzativi e analizza le performance, creando le condizioni per il miglioramento delle prestazioni del SGQ. L'altro ciclo, detto esterno, è quello relativo al rapporto tra organizzazione e cliente. Il cliente da un lato stabilisce i requisiti del prodotto/servizio aspettandosi di ottenerlo rispondente a quanto richiesto, e dall'altro fornisce indicazioni sul suo grado di soddisfazione; tali informazioni, associate a quelle acquisite internamente, permettono all'alta direzione di migliorare ed offrire prodotti/servizi sempre più rispondenti alle esigenze ed aspettative del cliente. L'intera ISO9001:2000 poggia, pertanto, sul cosiddetto ciclo "cliente-cliente". Infatti, partendo dall'ascolto delle esigenze del cliente, l'azienda deve trasformare tali bisogni in prodotti-servizi che una volta immessi sul mercato devono soddisfare le richieste del cliente stesso.

Fig 1.2 “Modello Di Processo del sistema di gestione per la qualità comune alle nuove norme ISO 9001:2001 e ISO



Quanto detto ci consente di introdurre altra novità della ISO/DIS 9001:2000, ovvero l'orientamento al cliente della gestione aziendale; a tal fine è importante sottolineare 2 concetti di cui si compone tale orientamento: il *customer focus* e la *customer satisfaction*¹. Per *customer focus* si deve intendere la capacità di comprendere i bisogni e le aspettative dei clienti, di tradurli in requisiti del prodotto/servizio e caratteristiche specifiche dell'offerta e di valutarne l'impatto sulla soddisfazione. Per misura della *customer satisfaction* (CS) si deve intendere sia l'ascolto puntuale del feedback dal cliente, sia la sua trasformazione in indicatori di performance dell'impresa per il miglioramento continuativo rispetto ai concorrenti.

Al *customer focus* tutta l'impresa si deve attenere in via prioritaria, mentre la misura della *customer satisfaction* costituisce l'obiettivo al cui raggiungimento l'impresa deve tendere.

¹ Meacci S. Valutazione dell'Impatto della ISO/DIS 9001 : 2000 sulla gestione del sistema di qualità aziendale i. Il caso "Databank". "U & C : " n. 6 giugno 2000, p.p.32

Altro aspetto di grande innovazione della ISO 9000:2000 è la grande importanza attribuita al concetto di Misurare. In questa ottica, infatti, il comitato di revisione da un lato ha ritenuto opportuno inserire tra gli 8 principi di buona gestione quello relativo alle *Decisioni basate sui dati di fatto*, che definisce decisioni efficaci quelle basate sull'analisi, logica ed intuitiva, di dati e informazioni reali, dall'altro ha incluso negli 8 paragrafi della ISO 9001:2000 quello "*Misurazioni, analisi e miglioramenti*" imponendo di fatto al fornitore di costruire un sistema per la misurazione, l'analisi e il miglioramento di elementi quali la soddisfazione dei clienti, l'andamento dei processi, la qualità di prodotti e l'efficacia delle verifiche ispettive. In particolare la CS, da traguardo da raggiungere, diventa un fatto sistemico, un *conditio sine qua non*, uno dei cardini della logica del miglioramento continuo oltre che una imprescindibile realtà della certificazione. In altri termini, il fornitore dovrà assicurare che tutte le aspettative dei clienti siano realmente ed adeguatamente considerate nel SGQ.

1.2 La Revisione Dello Standard e Le Principali Novità: ISO 9001:2008

Le ISO 9000 sono norme *volontarie* e *valide per tutti i settori produttivi*, sia industriali che di servizi. Nascono come strumento di tutela per il cliente, infatti, la loro principale funzione è "*garantire al cliente che l'organizzazione dell'azienda sia in grado di fornire quanto concordato, in maniera costante nel tempo*". Tuttavia, le norme ISO 9000, pur dichiarandosi valide per tutti i settori produttivi, tradiscono la loro origine industriale.

L'elaborazione di una Guida per i servizi (ISO 9004-2), rappresenta certamente un valido aiuto, ma non riesce a colmare tutte le difficoltà che le aziende di servizi incontrano nell'interpretare i requisiti della ISO 9001.²

Va sottolineata, in ogni caso, la finalità della normativa, che è quella di svolgere un preciso ruolo di guida-binario verso molteplici destinazioni. Tutte le attività imprenditoriali (industriali, commerciali, servizi, ecc.) interagiscono con l'ambiente e mirano a soddisfare le esigenze del cliente, il quale, pur avendo caratteristiche e connotazioni diversificate secondo le singole attività, ha alla base della sua soddisfazione la sensazione comune dell'ottimale rapporto Qualità-Prezzo dell'attività erogata ed acquisita.

Pertanto, la necessità di interpretare la Norma, adeguandone i contenuti secondo le esigenze dei settori più vari, costituisce probabilmente un punto di forza della stessa e non una carenza.

È stata pubblicata lo scorso 15 novembre 2008 la nuova versione della Norma ISO 9001, proprio con lo scopo di migliorare la comprensione e, conseguentemente, evitare che siano necessarie delle interpretazioni che possono portare e non applicare sempre nello stesso modo alcuni punti norma dello Standard e ricercare una sempre maggiore compatibilità con la Norma ISO 14001:2004 (che regola i Sistemi di Gestione Ambientale), in modo da facilitare ancora di più l'integrazione dei due strumenti. Venendo ai punti norma, i principali cambiamenti rispetto alla versione del 2000 riguardano i seguenti paragrafi:

Nel paragrafo "Sistema di Gestione per la qualità", nella versione del 2000 si specificava che l'organizzazione doveva identificare i processi, stabilire la sequenza e le interazioni fra essi e monitorare, misurare ed analizzare i processi. Inoltre, si parlava già dei processi che l'organizzazione dava all'esterno, affermando che questi dovevano essere controllati. Con la revisione, si chiarisce meglio come debbano essere

² **BARBARINO F.C., LEONARDI E.**, *Servizi di qualità - Modelli, norme ISO 9000 e strumenti operativi per progettare, gestire e migliorare il servizio nelle imprese, nel settore pubblico e nel non profit*. Cit., pp. 149-169.

posti sotto controllo da parte dell'organizzazione proprio quei processi che la stessa esternalizza.

Nel paragrafo dedicato alle "Risorse umane", nella versione del 2000 si affermava la necessità che il personale, le cui attività avevano influenza sulla qualità del prodotto, avesse un'adeguata istruzione, addestramento, abilità e preparazione. Con la revisione del 2008 si chiarisce meglio come debba essere ampliata la formazione del personale in quanto la conformità del prodotto ai requisiti è influenzata indirettamente dall'operato del personale stesso. Ciò si traduce in un maggior sforzo che le organizzazioni dovrebbero porre in essere al fine di istruire ed addestrare gli addetti.

Nel paragrafo dedicato all'"Ambiente di lavoro", la versione del 2000 affermava, genericamente, che l'organizzazione doveva definire e gestire le condizioni dell'ambiente di lavoro, in modo da assicurare la conformità ai requisiti dei prodotti. La revisione del 2008 viene in aiuto alle organizzazioni fornendo alcuni esempi su cosa si intenda per ambiente di lavoro e come l'organizzazione possa agire, in base alle proprie peculiarità.

Il paragrafo dei "Processi relativi al cliente" nella determinazione dei requisiti relativi al prodotto è stato puntualizzato nella versione del 2008 fornendo alcuni esempi su quali possano essere le attività successive alla consegna del prodotto al cliente o della erogazione di un servizio. In alcuni casi, tali azioni, che devono essere gestite all'interno del sistema qualità dell'organizzazione, possono accompagnare il cliente fino allo smaltimento del prodotto stesso. I requisiti relativi al prodotto e le azioni che ne derivano sono concetti importanti perché contribuiscono a definire correttamente il campo di applicazione del sistema.

Il paragrafo dedicato ai "Monitoraggi e misurazioni" prevede un sottoparagrafo dedicato alla valutazione della soddisfazione del cliente. La revisione del 2008 pone l'accento su come l'organizzazione possa

affrontare tale tema attraverso le informazioni che normalmente essa già possiede, quali, ad esempio, l'andamento delle vendite. In questo modo si invitano in maniera implicita le organizzazioni ad avvalersi di tali dati, ai quali si possono eventualmente affiancare le più classiche indagini di CS.

Per quanto concerne i temi delle "Azioni correttive" e delle "Azioni preventive", la nuova ISO 9001:2008 puntualizza che non sono tanto le azioni in sé a dover essere riesaminate dopo che sono state implementate, ma la loro effettiva efficacia nell'ambito del sistema di gestione.

1.3 Customer Satisfaction e Total Quality Management: Quale interazione?

Il TQM rappresenta un modello per la gestione totale della qualità esteso a tutte le attività aziendali, nessuna esclusa, che prevede il coinvolgimento di tutto il personale in un processo di miglioramento continuo.

Secondo KAORU ISHIKAWA, il TQM è un *“sistema per integrare le tecnologie della qualità in tutte le funzioni aziendali al fine di raggiungere la soddisfazione del cliente”*.²

Filosofia del TQM è guidare l'impresa verso l'ottenimento di risultati via via più significativi, secondo un'impostazione di lungo periodo, facendo leva sulla soddisfazione del cliente, sulla razionalizzazione dell'impiego delle risorse, sul miglioramento dell'efficacia e dell'efficienza dell'organizzazione e dei suoi processi, attraverso iniziative aventi per oggetto la globalità delle aree aziendali.³

Gli *elementi chiave* della Gestione Totale per la Qualità possono essere sintetizzati nei seguenti punti:

- ✓ focalizzazione sul cliente;
- ✓ sviluppo e coinvolgimento del personale;

² **BADIRU A.B., BABATUNDE J.A.**, *Practitioner's guide to Quality and Process Improvement*. Londra, Chapman & Hall, 1993.

³ **NEGRO G.**, *Organizzare la qualità nei servizi*. Milano, Il Sole 24 Ore, 1996.

- ✓ misurazione della qualità;
- ✓ miglioramento continuo.

1.3.1 Focalizzazione sul cliente

Una delle idee-forza della Qualità Totale è che *“Il cliente è al centro dell’organizzazione”*.⁴

Secondo il management classico, la direzione era al centro dell’organizzazione.

Dire che oggi il cliente diventa il centro dell’organizzazione non significa che la direzione ha perso il suo ruolo, ma solo che lo svolge in modo diverso: l’autorità della direzione dipende dalla sua capacità di trascinare tutta l’organizzazione a pensare nei termini del cliente e del concorrente, in modo creativo e responsabile.

Servire il cliente possibilmente meglio della concorrenza è un imperativo di sopravvivenza e sviluppo.⁵

E’ il cliente che giudica, non i tecnici e gli specialisti: solo la sua soddisfazione misura il valore aggiunto di un’impresa.

In questo contesto, la qualità potrebbe essere definita come *“il grado di soddisfazione del cliente al costo minimo e in misura comunque superiore alla concorrenza”*. La soddisfazione del cliente si ha solo quando le prestazioni dell’erogatore del servizio sono esattamente corrispondenti alle sue attese.⁶

Purtroppo, non sempre il cliente rende esplicite le sue aspettative, o perché le dà per scontate, o perché non è in grado di farlo; inoltre, se da un lato egli è molto esigente, dall’altro difficilmente manifesta la sua

⁴ **SALINI D.**, *Valutare la qualità nei servizi*. In *Qualità: cultura per il 2000 - Competizione globale e progresso civile*. XX Convegno nazionale dell’AICQ, atti del convegno, vol. C, Bologna, 15-17 maggio 2000, pp. 311-320.

⁵ **PORTANOVA V.**, *La soddisfazione del cliente obiettivo prioritario per il futuro*. In *Qualità: cultura per il 2000 - Competizione globale e progresso civile*. XX Convegno nazionale dell’AICQ, atti del convegno, vol. C, Bologna, 15-17 maggio 2000, pp. 209-213.

⁶ Ibidem.

eventuale insoddisfazione, ma semplicemente si rivolge altrove.⁷ Le strategie vincenti, dunque, si basano sempre più sulle grandi o piccole innovazioni che danno al cliente un margine in più di soddisfazione o di servizio (qualità, varietà, disponibilità...); ma in ogni caso, per dar frutti, la creatività dev'essere incentrata sulla soddisfazione del cliente, che è duratura solo al prezzo di una dinamica di incessante progresso.⁸

In Giappone, nelle aziende che applicano la Qualità Totale, i problemi sono definiti “tesori” e tutto il personale deve sapersi costruire tanti di questi tesori. Se in un'azienda, anziché trovarne di nuovi, ci si limita a risolvere solo i problemi che si hanno di fronte, si va verso la stagnazione. In assenza di questi “tesori” la tensione verso il miglioramento sarà debole.⁹

1.3.2 Sviluppo e coinvolgimento del personale

Il personale a contatto con il cliente ricopre una posizione difficile, perché deve erogare il servizio, personificare l'azienda agli occhi del cliente e cercare di soddisfarne le esigenze, difendendo contemporaneamente gli interessi economici dell'azienda.¹⁰

La gestione del personale è resa ancor più complessa dal gran numero di dipendenti che hanno contemporaneamente relazioni con lo stesso cliente; la numerosità di questi rapporti pone all'impresa il problema di mantenere standard di comportamento comuni a tutto il personale.

Il ruolo del fattore umano nei servizi è determinante, perché la qualità fornita al cliente e quindi la sua soddisfazione, sono soprattutto, il risultato del modo in cui il personale opera ed interagisce con esso.¹¹

⁷ CODA R., *Un controllo di gestione di “Qualità Totale”*. “Amministrazione & finanza”, anno XIII, n. 19, 15 ottobre 1998, pp. 33-37.

⁸ BASILICO M., CASTELLAZZI E., *Da oggi, per le aziende, professionisti di Qualità*. “Amministrazione & finanza”, anno XIV, n. 14, 31 luglio 1999, pp. 33-39.

⁹ GALGANO A., *Problemi come tesori*. “De Qualitate”, anno III, n. 1, gennaio 1994, pp. 16-20.

¹⁰ BERTOSSI G. ET AL., *Ruolo del personale e Customer Satisfaction*. “Sviluppo & Organizzazione”, n. 157, settembre/ottobre 1996, pp. 99-109.

¹¹ ALBRECHT K., *Al servizio del cliente interno ed esterno*. GRAMMA, ISEDI, 1992.

Ogni addetto deve essere messo in condizioni di saper gestire in autonomia i contatti con i clienti, in modo da tenere sempre sotto controllo la componente variabile del servizio, che nasce dal fatto che il cliente ha un ruolo di co-attore e possiede potenzialità di azione non sempre prevedibili.

Chi offre un servizio ha la consapevolezza che non sempre è possibile rendere “tangibile” ciò che viene fornito, dunque, la parte più “visi-bile” di un servizio sta proprio nelle persone che lo mettono in atto.

I comportamenti e gli atteggiamenti del personale rappresentano, dunque, più che un semplice “ingrediente” nel raggiungimento della soddisfazione. Affinché la componente relazionale del servizio corrisponda a quanto il cliente si aspetta, è necessario “investire” nel personale. Secondo quest’ottica, le persone che lavorano nella catena di attività, non devono essere considerate semplicemente come la “componente umana”, né il “fattore umano”, ma diventano una “risorsa” che nel tempo può contribuire in maniera determinante alla produttività ed alla redditività dell’azienda.¹²

Quest’approccio riguarda tutto il personale, sia il *front line* che ha diretto contatto con il pubblico, sia il *back office* che gestisce le attività nelle retrovie. Le due componenti hanno un ruolo ben diverso, ma sono fortemente legate. Il *front line*, infatti, è nelle condizioni di svolgere i suoi compiti, nel migliore dei modi e nel rispetto dei tempi, solo se ha alle spalle una valida attività di supporto del *back office*, in termini di disponibilità di prodotti, di cortesia, di comunicazioni verso il cliente, di gestione dei fornitori ecc.. Nel contempo, il *back office* ha continuo bisogno di verifiche sul suo operato in quanto, non avendo diretto contatto con il cliente, non ha conferma sul gradimento delle diverse componenti del servizio.¹³

¹² BARBARINO F.C., LEONARDI E., *La risorsa umana: chiave del servizio*. “De Qualitate”, anno VI, n. 3, marzo 1997, pp. 55-64.

¹³ ROSANDER A.C., *Vivere per il cliente*. Editoriale Itaca, 1994.

1.3.3 Misurazione della qualità

Per comprendere in modo oggettivo i risultati prodotti dai passi verso una migliore qualità e quindi , occorrono indicatori, misure quantitative: la misurazione della qualità può prendere spunto dai fattori critici di successo, ovvero da quelle caratteristiche di un reparto che il cliente considera più importanti.

Sta di fatto tuttavia che alcuni fattori critici non sono facilmente ed efficacemente quantificabili.¹⁴

1.3.4 Miglioramento continuo

Tutti gli strumenti della qualità totale, in definitiva, puntano alla soddisfazione del cliente.

Aspetto tipico della natura umana è quello di non essere mai soddisfatti a sufficienza e questo vale anche per il cliente: quando ha ottenuto quello che desiderava già pensa a qualcosa di meglio, mentre i concorrenti stanno già lavorando per offrirglielo. Ecco, quindi, la necessità di non fermarsi mai nei processi di miglioramento, nonostante la certificazione della qualità sia già stata rilasciata.

In conclusione, la qualità del servizio non deve essere intesa come un traguardo che si raggiunge in maniera definitiva, ma un fine verso cui tendere.¹⁵

¹⁴ **PIERUCCI D.**, *Normazione tecnica e certificazione di qualità, un approccio integrato alla soddisfazione del cliente*. In *Qualità: cultura per il 2000 - Competizione globale e progresso civile*. XX Convegno nazionale dell'AICQ, atti del convegno, vol. A, Bologna, 15-17 maggio 2000, pp. 153-158.

¹⁵ **BARBARINO F.C., LEONARDI E.**, *Servizi di qualità - Modelli, norme ISO 9000 e strumenti operativi per progettare, gestire e migliorare il servizio nelle imprese, nel settore pubblico e nel non profit*. Milano, Il Sole 24 Ore, 1997, pp. 65-68

1.4 La Customer Satisfaction come strumento gestionale

Lo sviluppo cronologico delle relazioni Cliente/Azienda può essere semplificato in quattro fasi:

1) Alla partenza il cliente ha le sue aspettative, “Qualità attesa”. L’azienda, grazie al suo processo di ascolto, trasforma tali aspettative in “qualità desiderata”;

2) La seconda tappa consiste nel passare dalla qualità desiderata alla “qualità realizzata”;

3) Una volta realizzata tale qualità, l’azienda deve trasmetterlo al cliente attraverso il processo di comunicazione/vendita che consente al cliente, dopo l’acquisto, di determinare la sua percezione (qualità percepita);

4) La comparazione tra la qualità attesa (prima dell’acquisto) e la qualità percepita (dopo l’acquisto) genera la CS.

Si può notare che la formazione di tale CS ha delle ripercussioni sulle attitudini future del cliente nei riguardi del fornitore: in effetti la qualità percepita lo informa su quello che può realmente ottenere modificando, di conseguenza, le sue future aspettative.

Quindi, attraverso tale schematizzazione la soddisfazione è vista come il grado di adeguazione tra l’inizio e la fine di tale processo.

L’insoddisfazione può, dunque, essere generata da tre tipi di scarti negativi:

- . scarto tra qualità attesa e qualità realizzata.

Tale scarto può essere generato o da errori durante la fase di ascolto o da errori nella trasformazione delle aspettative (bisogni) del cliente in caratteristiche del prodotto/servizio da realizzare;

- . scarto tra qualità desiderata e qualità realizzata.

In questo caso la risoluzione di tale problema va ricercata nel processo di realizzazione del prodotto/servizio;

- . scarto tra qualità realizzata e qualità percepita.

In questo caso è il processo di comunicazione o vendita dell'azienda che non funziona come dovrebbe.

1.4.1 L'analisi della Customer Satisfaction nell'ambito del processo di "Ascolto del cliente"

"Il cliente prima di tutto", "Inserire il cliente nel cuore dei processi dell'azienda"... questi e tanti altri sono gli slogan che normalmente sono utilizzati per ricordare ai dipendenti ed ai dirigenti che i clienti devono costituire una priorità per l'azienda. Per perseguire tale principio risulta fondamentale ascoltare il cliente.

Esistono diversi modi per ascoltare il cliente, il migliore è sicuramente quello di garantire una presenza quotidiana presso il cliente stesso. Infatti, tutte le indagini e le migliori metodologie di analisi non possono mai eguagliare l'importanza di un ascolto attivo sul campo giorno dopo giorno. Tale sistema presenta, tuttavia, alcuni inconvenienti: anzitutto è molto dispendioso per l'azienda dover garantire una presenza quotidiana presso ciascun cliente; inoltre, non sempre le informazioni raccolte dal front-office risalgono in tempi brevi fino al management; in ultimo, non sempre i clienti esternano le loro insoddisfazioni. L'insieme di tutto ciò limita fortemente la capacità del management di prendere decisioni a fronte di tale forma di ascolto.

Le altre forme di ascolto sono (fig.1.1)

- Il sistema di gestione dei reclami;
- Gli studi inerenti "i Clienti Perduti";
- Gli studi del tipo "Clienti Misteriosi"
- Gli studi sulla "Misurazione della Customer Satisfaction"

In questo contesto, le inchieste volte ad analizzare la CS non costituiscono che una pietra dell'edificio, certamente indispensabile, ma insufficiente per gestire efficacemente il "patrimonio Cliente".

Fig 1.1



1.4.2 I reclami dei clienti

Un reale ascolto del cliente sul campo favorisce la manifestazione dei motivi di insoddisfazione dei clienti ed è compito dell'organizzazione dell'azienda farli risalire quanto più rapidamente possibile fino al management. Tali reclami costituiscono una vera miniera d'oro purché siano ascoltate e prese in considerazione per il miglioramento.

A questo punto è opportuno focalizzarsi sugli strumenti per individuare un cliente insoddisfatto e comprenderne i motivi.

Per individuare i clienti insoddisfatti esistono due sistemi:

- . effettuare un studio di soddisfazione della clientela;
- . raccogliere i reclami e le lamentele.

Quest'ultima soluzione presenta il vantaggio di individuare immediatamente, rispetto al primo metodo che invece necessita di uno studio accurato, i clienti insoddisfatti ed i motivi dell'insoddisfazione; il limite di tale sistema è che i clienti raramente formalizzano le loro lamentele poiché lo ritengono privo di utilità. Dunque, analizzare i reclami

significa individuare solo una parte dei clienti insoddisfatti ovvero quelli che rappresentano la punta dell'iceberg dell'insieme degli insoddisfatti.

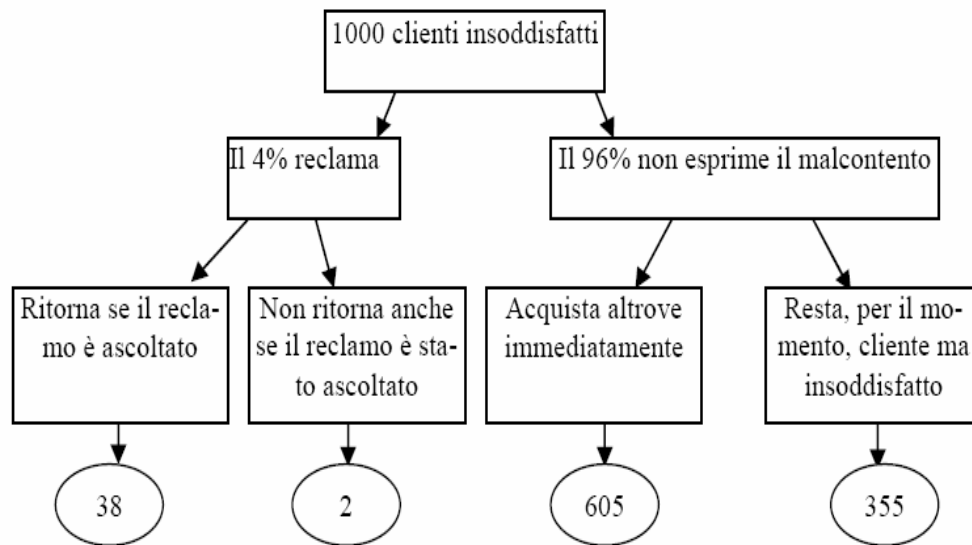


Figura 1.2: L'incidenza dei reclami (Fonte Nielsen)

Da ciò appare evidente che raccogliere i reclami, analizzarli e soprattutto individuare una possibile soluzione è un'attività di fondamentale importanza nell'ottica del miglioramento delle prestazioni. Tuttavia, limitarsi a questo strumento di ascolto significa risolvere solo una ridotta percentuale dei problemi degli insoddisfatti. Infatti, secondo i dati provenienti uno studio effettuato dalla Nielsen nel 1999, solo il 4% dei clienti insoddisfatti formalizza un reclamo (fig.1.2).

1.4.3 Analisi dei clienti perduti

Per un'azienda una risorsa importante di miglioramento in termini di soddisfazione e di retention dei clienti consiste nel comprendere due fenomeni complementari:

- Quali sono i motivi di soddisfazione più importanti che la fanno preferire ai concorrenti;
- Perché alcuni clienti la abbandonano.

Lo studio dei clienti perduti riguarda questo secondo aspetto e il suo scopo deve essere di:

- Comprendere le logiche di acquisto dei clienti persi;
- Esplicare gli elementi che hanno determinato l'abbandono;
- Inserire tali elementi nel processo aziendale di erogazione del servizio e/o prodotto;
- individuare le azioni correttive e preventive da adottare con i relativi responsabili ed esecutori.

1.4.3. Le indagini sui clienti misteriosi

La percezione dei clienti, e dunque la loro soddisfazione dipende dalla qualità oggettiva delle prestazioni fornite. Questa qualità oggettiva è più facilmente misurabile quando si è nel settore della produzione di beni tangibili (ritardi, non conformità ecc.) rispetto al settore dei servizi.

L'obiettivo di tale studio è di verificare la conformità alle specifiche stabilite in chiave progettuale del prodotto-servizio erogato. Lo sviluppo di tale analisi si ottiene in due fasi: inizialmente è necessari individuare i parametri del prodotto o servizio da valutare con relative unità di misura, successivamente, attraverso dei clienti fittizi fruire del prodotto servizi e rilevare i valori assunti dai parametri in esame.

1.3 Il processo di formazione della Customer Satisfaction

Il processo di formazione della soddisfazione si basa su un modello di comportamento del consumatore chiamato multi-attributo. Questo modello che costituisce uno dei pilastri della teoria classica del marketing, discende dall'analisi microeconomica del processo di acquisto. Al momento di una situazione di acquisto, il consumatore razionale effettua la sua scelta cercando di massimizzare il "beneficio del consumatore". Per questa ragione l'acquirente deve essere in grado di valutare (ovvero di rendere

misurabile) il beneficio potenziale di ciascuno dei prodotti/servizi in concorrenza, al fine di poter scegliere razionalmente quello che gli genera il più grande “beneficio percepito”. Il consumatore razionale considera, dunque, il prodotto non più come unità ma come la somma di benefici potenziali. Se indichiamo con il termine di attributo ciascuna di queste caratteristiche o dimensioni, ogni prodotto/servizio diviene dunque un paniere di attributi. Il consumatore razionale cercherà, evidentemente, di riempire al massimo il suo paniere al fine di ottimizzare il suo “beneficio”.

Il modello multi-attributo propone dunque di adottare un ragionamento secondo il quale ogni attributo contribuisce alla determinazione della soddisfazione globale. In tale ottica appare naturale definire la CS come una grandezza multidimensionale. In tale ottica appare evidente che la CS condiziona fortemente le scelte di acquisto del consumatore e, di conseguenza, la competitività delle imprese produttrici. Pertanto la possibilità per le aziende di poter disporre di misure accurate sulla valutazione dell'utilizzatore del prodotto/servizio diviene altrettanto importante del know-how tecnologico alla base del sistema produttivo o di erogazione. In questo contesto il contributo che può dare la metodologia statistica per la messa a punto di misurazioni accurate, di modelli interpretativi e di analisi rigorose della CS appare rilevante. Si sta delineando, in effetti, un settore in cui la statistica potrà giocare un ruolo innovativo e determinante, con risultati di grande utilità applicativa.

1.6 Le tre principali caratteristiche della soddisfazione: soggettività, relatività ed evoluzione

Il modello di valutazione di un prodotto/servizio da parte di un cliente si basa su tre criteri principali: la soggettività, la relatività e l'evoluzione. La soddisfazione del cliente dipende dalla sua percezione del prodotto/servizio e non dalla realtà (principio della soggettività). Oltre a dipendere dalla percezione del cliente, la soddisfazione varia anche in relazione al livello

delle aspettative. Come possono 2 clienti utilizzare lo stesso prodotto/servizio nelle stesse condizioni ed avere opinioni radicalmente differenti?

Semplicemente perché hanno nei confronti del prodotto/servizio aspettative differenti (principio della relatività). Di conseguenza è possibile sottolineare come il ruolo della segmentazione nella gestione dei clienti è di grande importanza, in quanto attraverso essa è possibile identificare dei gruppi omogenei di clienti aventi aspettative simili, in modo da creare per ciascun gruppo offerte adatte. In ultimo la soddisfazione si modifica in relazione al tempo a due livelli distinti: in funzione della evoluzione delle attese e degli standard esistenti ed in funzione del ciclo di utilizzazione del prodotto/servizio (principio della evolutività).

Per quanto concerne il primo aspetto, si può affermare che, così come il mercato concorrenziale “costringe” i fornitori a migliorare le proprie performance relative allo scopo di essere preferiti, parimenti, migliorano i livelli medi delle offerte e dunque gli standard di riferimento. L’evoluzione di tali standard genera un’evoluzione delle aspettative dei clienti e un conseguente mutamento del processo di formazione della CS.

In riferimento al secondo aspetto è noto che la misura della CS è realizzata in un dato momento e che tale misura si modifica durante il periodo di utilizzazione del prodotto/servizio. In effetti, immediatamente dopo l’acquisto (il momento generalmente più utilizzato per le indagini di CS) la curva di soddisfazione raggiunge il suo picco più alto per poi decrescere nel tempo. Questo fenomeno può essere giustificato in due modi:

- 1) La misura della soddisfazione subito dopo l’acquisto è influenzata dal cosiddetto effetto della “disonia cognitiva” ovvero lo stato transitorio che colpisce il consumatore quando il suo comportamento si allinea con le sue attitudini;

- 2) Il miglioramento continuo delle offerte sui mercati, di fatto, banalizza ciò che all’istante dell’acquisto rappresentava un vantaggio concorrenziale e dunque una sorgente di soddisfazione per il cliente.

Si noti che in questo caso non è l'evoluzione delle aspettative che fa diminuire la soddisfazione ma è l'obsolescenza del prodotto/servizio acquistato in rapporto ai nuovi livelli del mercato.

Da un punto di vista economico e sociale, i servizi definiscono un'area produttiva caratterizzata da un'eterogeneità molto marcata. Numerosi sono stati i tentativi in letteratura di classificare i servizi ricorrendo a parametri, fra loro assai diversi, presenti in maniera più o meno incisiva: intensità di lavoro, componente tecnologica, grado di personalizzazione, grado di interazione tra fornitore e cliente, presenza di supporto tangibile, criticità del rapporto con il personale ecc.. Ma i servizi costituiscono un mondo così vario ed eterogeneo che può risultare molto limitativo inquadralo in rigide categorie analitiche. Ci sono però alcuni elementi che sono comuni e ricorrenti, in misura più o meno marcata, in tutti i servizi.

Il servizio rappresenta “un insieme di attività”, assistite o meno da prodotti, che nasce come risposta alle diverse esigenze del cliente. Maggiore sarà la capacità di conoscere e comprendere i bisogni ed i desideri del cliente, maggiore sarà la probabilità di fornirgli le giuste risposte nei tempi e nei modi più idonei.

Potremmo affermare, in linea teorica, che la produzione/erogazione di un servizio nasce dalla necessità che qualcuno si preoccupi di fornire ad un terzo la possibilità di disporre, accedere o utilizzare fonti e risorse in cambio di un corrispettivo per il servizio reso.

In prima istanza, dunque, la nascita dei servizi va ricercata in due momenti simultanei:

- la necessità di soddisfare un bisogno;
- l'individuazione di un bisogno esistente da soddisfare.

Il servizio rappresenta un'attività di problem solving del cliente, quindi nasce e si deve sviluppare facendo leva su una forte interazione fra utente ed erogatore del servizio stesso.

La capacità di produrre un servizio parte dalla conoscenza del cliente.

Nei diversi tipi di servizio c'è la comune esigenza della risoluzione di un “problema”, in cui la componente di individualità è a volte

determinante, il che rende più complesso il servizio in termini di aspettative e soddisfazione del cliente. Sia nel servizio associato a prodotti che in quello puro, cioè senza nessun supporto tangibile, è fondamentale che vi sia integrazione fra le due aree del personale indicate come front line e back office.

All'interno della Grande Distribuzione (Gdo) capire a fondo le aspettative della clientela e fare di tutto per soddisfarle è il miglior modo per assicurarsi il successo, perché un cliente davvero soddisfatto:

- tende ad essere fedele ;
- spesso diventa un vero e proprio promotore dell'impresa;
- è relativamente meno sensibile ai differenziali di prezzo;
- consente di avere un più elevato livello di redditività

E' chiaro che considerando la realtà economica attuale, caratterizzata da una restrizione economica, ma soprattutto da concorrenti molto agguerriti, obiettivo del responsabile del punto di vendita, è innanzitutto quello di riuscire captare prima dei concorrenti quali sono le esigenze dei clienti, imparare ad ascoltarli e cercare di offrire loro, con competenza ed efficienza un servizio sempre differente rispetto alla concorrenza, anche perché un cliente soddisfatto non solo tornerà di nuovo all'interno del ns supermercato, ma ci fa da passaparola positivo. Il responsabile però deve essere altresì in grado non solo di offrire un certo servizio, ma di garantirlo nel tempo, con la massima trasparenza e professionalità.

Da uno studio Università - Responsabile della qualità di una tra le più grandi catene distributive presenti attualmente sul mercato, al fine di monitorare l'andamento dei clienti all'interno delle loro strutture si è convenuto nel focalizzare l'attenzione solo su alcuni aspetti, definiti indispensabili al fine di soddisfare e fidelizzare sempre più un numero maggiore di clienti così come evidenziato nella seguente tabella:

VARIABILI	OBIETTIVO
Accessibilità	Garantire l'accesso al servizio dell'utente in termini di segnaletica stradale e area parcheggio
Efficacia	Capacità di rispondere alle necessità dei clienti
Affidabilità	Garanzia di continuità del servizio e di Sicurezza nel suo utilizzo
Flessibilità	Capacità di andare incontro alle specifiche esigenze dei clienti
Trasparenza	Chiarezza sulle caratteristiche del servizio e sulle condizioni di utilizzo
Professionalità	Competenza ed efficienza con cui il servizio è reso
Gradevolezza	Gradevolezza del personale addetto. Aspetto e pulizia degli ambienti
Convenienza	Rapporto tra i benefici dati dal servizio ed i suoi costi
Ascolto	Capacità di ascoltare e risolvere i problemi dei clienti

Oggigiorno la grande distribuzione sta focalizzando la sua attenzione sempre più verso un monitoraggio quotidiano della propria clientela, grazie soprattutto all'analisi delle fidelity card e quindi dei comportamenti dei propri clienti, con l'unico obiettivo che è quello di fidelizzare un numero sempre maggiore di persone.

Tale monitoraggio focalizzando particolare attenzione sugli aspetti sopra evidenziati ha come obiettivo quello di apportare informazioni circa il livello di soddisfazione complessivo percepito dai clienti, fornire notizie utili al management qualora ci dovessero essere azioni correttive da intraprendere, far sì che tali dati risultano essere equiparabili nel tempo, questo grazie al fatto di somministrare questionari standardizzati.

Soddisfare il cliente significa anche coinvolgere il personale dipendente di tutti i reparti, in modo tale da avere un impiego più efficiente ed

efficace di tutte le risorse disponibili, con l'obiettivo di migliorare la qualità dei servizi offerti.

1.7 Analisi delle Fidelity card

Uno dei principali strumenti di fidelizzazione nel mercato della grande distribuzione è senza dubbio la carta fedeltà, cioè quella tessera formato bancomat attraverso la quale i supermercati e i grandi magazzini da un lato offrono la possibilità ai propri clienti di accedere a campagne promozionali o di partecipare a raccolte punti e dall'altro si assicurano il formidabile vantaggio strategico di riuscire a dare un nome e un volto a un anonimo scontrino.

Il meccanismo prende le mosse dal codice a barre che sono un insieme di elementi grafici a contrasto elevato disposti in modo da poter essere facilmente letti da un sensore e decodificati tramite un apposito circuito integrato. Il codice a barra di ogni prodotto acquistato e il codice identificativo della carta vengono rilevati dal lettore ottico posizionato alle casse. Un apposito data base raccoglie queste informazioni e li mette a disposizione dei responsabili marketing che li possono utilizzare per definire le future azioni commerciali. Grazie alla tessera fedeltà ogni singolo cliente diventa un consumatore dai contorni ben delineati. Di lui si conosce il suo comportamento : cosa compra, quanto spende in media ogni mese, la fascia oraria della giornata in cui effettua gli acquisti. È ovvio, però, che i consumatori devono lasciarsi coinvolgere: non basta che essi accettino di diventare possessori di una carta ma devono anche costantemente utilizzarla.

I dati forniti dalle carte fedeltà sul comportamento di acquisto della clientela costituiscono un importante patrimonio informativo che l'azienda, attraverso le più moderne tecnologie di Business Intelligence, può utilizzare e trasformare in "sapere". Le soluzioni di data mining consentono proprio questo processo di estrazione di conoscenza da banche dati di grandi dimensioni tramite l'applicazione di particolari algoritmi che

individuano le relazioni latenti tra le informazioni e le fanno emergere . È indubbio che, per un'azienda, può risultare di fondamentale importanza riuscire ad esempio a individuare i clienti potenzialmente a rischio di churn.

Questo permette di attivare nei loro confronti mirate azioni di fidelizzazione. In tal senso i dati relativi alla clientela possono essere elaborati e interpretati utilizzando una particolare metodologia statistica quale l'analisi della sopravvivenza che permette di studiare l'evoluzione nel tempo di un preciso fenomeno a partire dalla relazione che intercorre tra il fenomeno stesso e le caratteristiche proprie delle unità o del contesto esaminati.

CAPITOLO SECONDO

Modelli Per La Misurazione Della Customer Satisfaction

Sommario: **2.1** Misurare la Customer Satisfaction - **2.2** Metodologie utilizzate per la valutazione della soddisfazione nella Grande distribuzione - **2.2.1** Nei modelli diretti formativi o compositivi - **2.2.2** Modelli Diretti Esplicativi o Decompositivi - **2.2.3** Modelli Di Reti Bayesiane - **2.2.4** Modelli Strutturali

2.1 Misurare la Customer Satisfaction

La CS può essere definita come il risultato del confronto da parte del cliente dei costi e dei benefici derivanti dall'acquisto e utilizzo del prodotto/servizio in relazione alle aspettative. Il cliente sarà "soddisfatto" se i benefici dell'acquisto e dell'utilizzo sono conformi alle sue aspettative, ovvero se i benefici sono superiori alle sue aspettative, relativamente al costo sostenuto per acquistare il prodotto/servizio.¹

La CS è teorizzabile come una reazione emotiva del cliente, in quanto è legata al rapporto tra l'aspettativa del prodotto/servizio e l'"immagine" scaturita dall'esperienza d'uso. In particolare, nella grande distribuzione, PARASURAMAN, ZEITHAML e BARRY² sostengono che la soddisfazione o l'insoddisfazione del cliente è determinata dal confronto che egli opera tra la percezione del servizio fruito e le aspettative sullo stesso.

¹ BERTOSSI G. ET AL, *Ruolo del personale e Customer Satisfaction*. Cit., pp. 99-109.

² PARASURAMAN A., ZEITHAML V.A., BERRY L.L., *A conceptual model of service quality and its implications for future research*. "Journal of Marketing", vol. 49, 1985, pp. 41-50.

In altri termini, la soddisfazione si configura come una percezione della clientela su cosa sia stato offerto dall'azienda e sul come (attraverso quali prassi) gli sia stato fornito. Molto spesso, tuttavia, si tende a confondere il concetto di CS con quello di qualità, essendo tali concetti strettamente collegati ed "astratti".

In generale, mentre la qualità dei prodotti/servizi può essere intesa sia in modo oggettivo (vale a dire, come conformità alle specifiche di progettazione) che in modo soggettivo (cioè, come capacità di soddisfare le esigenze dei consumatori), la CS può essere misurata solo in modo soggettivo. Secondo CRONIN e TAYLOR,³ in riferimento alla qualità nei servizi, la CS è un antecedente della qualità di servizio percepita. Inoltre, mentre la percezione della qualità dell'offerta aziendale si forma nel lungo termine, la CS viene espressa su un singolo prodotto/servizio. La distinzione tra CS e qualità, elaborata soprattutto a livello teorico, comporta per l'azienda la necessità sia di soddisfare pienamente il cliente, sia di realizzare il massimo grado di qualità percepita.⁴

Secondo uno studio sperimentale, teorizzato in un modello esplicativo meglio noto come modello della "Attractive Quality" o "Modello di Kano",⁵ dal nome del suo autore, è stato evidenziato che gli elementi costituenti la soddisfazione possono essere catalogati secondo una gerarchia, con accentuate e importanti conseguenze sulla progettazione e produzione del prodotto/servizio.

Stando al succitato modello di esplicitazione, i fattori che determinano la soddisfazione possono distinguersi in tre classi, ognuna delle quali contribuisce in maniera peculiare alla soddisfazione complessiva del cliente:

³ CRONIN J., TAYLOR S.A., *Measuring service quality: a reexamination and extension*. "Journal of marketing", vol. 56, luglio 1992, pp. 55-68.

⁴ VALDANI E., BUSACCA B., *La Customer Satisfaction: specificità, analisi e management*. "Micro & macro marketing", anno IV, n. 3, dicembre 1995, pp. 315-343.

⁵ Si veda in proposito: KANO N., *Business strategies for the 21st century and attractive quality creation, proceeding of international conference for quality*. Yokoama, 1996, pp. 105-108; ARATO G., *Come conquistare il cliente. Il Modello di Kano e la creazione della "Attractive Quality"*. "Qualità", anno XXVIII, n. 5, settembre/ottobre 1998, pp. 17-19; HINTERHUBER H. ET AL, *Un modello semiquantitativo per la valutazione della soddisfazione del cliente*. "Micro & macro marketing", anno VI, n. 1, aprile 1997, pp. 127-143.

- i *fattori di base* rappresentano una specie di condizione indispensabile, ma non sufficiente al conseguimento dell'elevato grado di soddisfazione dei clienti. In altre parole, se essi non sono adeguati, il cliente resterà insoddisfatto; al contrario, se essi sono appropriati, non necessariamente conferiranno da soli una soddisfazione elevata;
- i *fattori prestazionali*, legati alla *performance* del prodotto/servizio, incidono in modo direttamente proporzionale sulla soddisfazione della clientela, pur con un'influenza contenuta. In altri termini, se essi sono adeguati, determineranno un aumento della soddisfazione, a condizione sempre che vi sia anche un'adeguatezza dei fattori di base;
- i *fattori di "delightment"*, che "deliziano e diletano" il cliente, sono quelli che spingono significativamente la soddisfazione verso gradi più alti ed agiscono positivamente sulla tendenza alla fedeltà o alla selezione dei fornitori. In presenza di fattori di base e prestazionali adeguati, fattori di "delightment" positivi sorprendono piacevolmente il cliente e ne incrementano notevolmente la sua soddisfazione complessiva.

Quest'ultima classe di fattori è quella che, per Kano, determina la creazione della "attractive quality".

L'analisi approfondita e metodica della soddisfazione del cliente rappresenta per l'azienda l'atto conoscitivo peculiare per il raggiungimento ed il mantenimento di una leadership competitiva.⁶ In particolare, le indagini della CS rappresentano il riferimento ed il sostegno a tutte le attività aziendali, in special modo nei seguenti principali settori:

- ✓ riconoscimento ed incremento delle caratteristiche dei prodotti/servizi offerti;
- ✓ definizione e controllo dei modi di realizzazione dei prodotti/servizi;
- ✓ compartecipazione del personale che contatta i clienti;
- ✓ sostegno alla pubblicità ed alla comunicazione aziendale;
- ✓ supporto alla direzione nelle scelte strategiche.

⁶ **ALTIERI L.**, *Servizio qualità efficienza sono in antitesi?*. "Logistica Management", n. 90, novembre 1998, pp. 65-67.

Il contatto e l'ascolto dei clienti (*orientamento all'Audit*) rappresentano elementi determinanti per l'individuazione delle caratteristiche di nuovi prodotti/servizi ed implicano la possibilità d'individuazione di nuove aree di espansione commerciale. Un'analisi dettagliata della CS consente di rivedere e controllare i processi di produzione/erogazione, adeguandoli alle molteplici e mutevoli necessità dei clienti. L'ascolto può, oltre a contribuire all'acquisizione di utili informazioni nella progettazione delle modalità di erogazione dell'offerta aziendale, consentire anche di realizzare prodotti/servizi secondo il grado di soddisfazione voluto dal cliente fin dall'inizio, evitando aggravii per eventuali correttivi.

I sondaggi di CS concorrono concretamente alla divulgazione e all'espansione della mentalità commerciale del cliente, propiziando nell'azienda un solido e proficuo orientamento verso il cliente da parte di tutto il personale, in special modo quello di contatto. Inoltre, l'individuazione e la registrazione delle richieste del cliente possono rafforzare nei vertici aziendali la consapevolezza della necessità di migliorare i livelli professionali delle categorie del personale.

I rilievi delle esigenze del cliente permettono anche di individuare le problematiche su cui basare la pubblicità e, in generale, la comunicazione aziendale. L'analisi dei dati forniti dalla clientela, dà, infatti, la possibilità all'azienda di orchestrare adeguatamente la comunicazione, orientandola ed applicandola sui problemi più impellenti per la clientela e nei posti in cui l'azienda potrà conquistare un ruolo proficuo.

L'analisi inerente alla soddisfazione del cliente consente al vertice aziendale di determinare le possibili aree in cui potranno essere lesi gli interessi aziendali. Quest'analisi consente inoltre alla direzione di avere una panoramica delle potenzialità dell'offerta aziendale nella sua poliedricità e, di conseguenza, favorisce l'adozione di opportune strategie.

Nonostante l'importanza strategica assunta dalla CS, sono ancora oggi poche le aziende italiane che pongono il cliente al centro del loro sistema di produzione/erogazione e che perseguono l'obiettivo della sua

soddisfazione.⁷ In genere, le aziende di medie/grandi dimensioni presentano una maggiore propensione verso la CS, ma molte di esse, sia pubbliche che private, hanno finora affrontato questa tematica come se il loro obiettivo si esaurisse nel perseguire la soddisfazione del cliente. In realtà, l'obiettivo dell'azienda non è solo quello di soddisfare i clienti, bensì, di poter influire sui loro comportamenti d'uso e consumo del prodotto/servizio, attraverso un'accresciuta soddisfazione.

La CS diventa, pertanto, per ogni azienda, il fulcro propulsore nell'adozione delle strategie e dei programmi intenti al continuo miglioramento della qualità, sia tecnica che funzionale.⁸

Compito dell'azienda è, in conclusione, allocare in modo ottimale le proprie risorse al fine di massimizzare il rapporto tra gli sforzi effettuati per migliorare il servizio ed i benefici che ne derivano, in termini di fidelizzazione, maggiori acquisti, passaparola e, in ultima analisi, di risultati aziendali.⁹

Appare indiscutibile che la CS sia il fine ultimo e primario dell'attività delle imprese: ciò che può costituire uno spunto di riflessione sono gli strumenti e le modalità attraverso le quali perseguire tale obiettivo.

Alcuni autori¹⁰ sostengono che potrebbe essere concettualmente poco corretto raggiungere la CS ponendo il cliente al centro dell'attività delle organizzazioni, in quanto le opportunità di successo di un'impresa risiedono essenzialmente nella sua capacità innovativa, dunque andrebbe rilanciato il ruolo propulsivo dell'imprenditore.

Le innovazioni rappresentano l'anima del progresso, e solo una costante attività innovativa consente alle imprese di mantenere ed accrescere il proprio mercato.

⁷ **TRONCONI A.**, *Il controllo di business nell'impresa di servizi*. "De Qualitate", anno V, n. 6, giugno 1996, pp. 55-64.

⁸ **SCARPINATO M.**, *La valutazione della Customer Satisfaction come strumento di analisi strategica*. "Sviluppo & Organizzazione", n. 141, gennaio/febbraio 1994, pp. 2-6.

⁹ **GUALTIERI F., POTIÉ C.**, *Dall'assicurazione della qualità alla fidelizzazione del cliente*. "De Qualitate", anno X, n. 5, maggio 2000, pp. 19-31.

¹⁰ **CAPPELLI L., RICCIO A.V.**, *Customer satisfaction e imprenditore schumpeteriano: conflitti e convergenze*. In *Qualità: cultura per il 2000 - Competizione globale e progresso civile*. XX Convegno nazionale dell'AICQ, atti del convegno, vol. C, Bologna, 15-17 maggio 2000, pp. 250-264.

I sostenitori della tesi della centralità del cliente¹¹ ritengono che il processo innovativo debba essere “trainato” dai consumatori; secondo questo principio, l’obiettivo della CS si può raggiungere solo attraverso una “rivoluzione” concettuale, che ponga idealmente il cliente al timone dell’azienda, assegnandogli il compito di dettare le regole del gioco.

Problema di fondo è che, un’impresa che rinunci al ruolo “attivo” dell’imprenditore, rinuncia di fatto all’attività che ne identifica la stessa esistenza: quella di produrre innovazioni.

La condivisione del principio secondo cui l’azienda deve essere “*customer driven*”, ossia guidata dal cliente, può portare ad un livellamento progressivo delle condizioni dell’offerta, volta al puro inseguimento del cliente. Ma il cliente è, per la sua stessa natura, concettualmente “ignorante”: egli è infatti *consumatore* e non possiede le caratteristiche proprie di chi riveste la funzione imprenditoriale.

Secondo l’approccio *schumpeteriano*,¹² invece, che individua la ragione del cambiamento nell’attività innovativa, viene ristabilito il corretto ruolo di ciascuna parte sul mercato: il cliente è il consumatore, e come tale è esterno ed estraneo all’impresa, mentre è l’imprenditore ad assumere la funzione cardine dell’impresa, quella di realizzare innovazioni. Queste ultime, rappresentano l’unica fonte di vantaggio competitivo per l’azienda.

Anche il concetto di qualità, se non inserito in quello più ampio di innovazione, appare privo di un concreto significato.

L’obiettivo della CS deve essere perseguito, di certo, avendo come riferimento ultimo il cliente, ma non mettendolo al centro dell’impresa, affidandogli compiti che non gli sono propri e per i quali egli non ha alcuna competenza.

Coerentemente con l’impostazione schumpeteriana, il raggiungimento della CS deve prevedere un modello aziendale che rilanci la centralità

¹¹ CHIACCHIERINI E., *Tecnologia e produzione*. ed. Kappa, Roma, 1996.

¹² SCHUMPETER J.A., *The theory of Economic Development*. Harvard University Press, Cambridge, 1934.

dell'imprenditore e quella dell'innovazione come unico strumento per perseguire la soddisfazione del cliente.

2.2 Metodologie utilizzate per la valutazione della soddisfazione nella Grande distribuzione

Nell'intento di fornire una chiave di interpretazione dei meccanismi di erogazione e di definire degli standard di riferimento, il servizio è diventato oggetto di studi sempre più accurati. Parallelamente, si è affermata la necessità di mettere a punto nuovi strumenti in grado di valutarne la qualità.

Spesso, l'insoddisfazione di un cliente si traduce, per il punto vendita, nella perdita del cliente stesso. L'esperienza degli operatori del settore, generalmente, non costituisce un metodo oggettivo d'indagine della qualità di un servizio. Ciò si traduce nell'impossibilità di avere un quadro preciso della situazione e, soprattutto, nella mancanza di indicazioni sugli aspetti del servizio che hanno generato un processo di erogazione non soddisfacente.

I metodi tradizionali per il controllo in linea dei processi manifatturieri, talvolta, si rivelano inadeguati ad essere utilizzati nel contesto della Gdo. Le ragioni sono diverse: da una parte, il peso rilevante del fattore umano, dall'altra, la difficoltà di gestire contemporaneamente variabili oggettive, soggettive, relazionali ed organizzative.

Gli aspetti principali che accomunano gli strumenti per effettuare valutazioni della qualità dei servizi sono:

- ✓ l'utilizzo di questionari;
- ✓ il riconoscimento del carattere multidimensionale della qualità;
- ✓ il prendere in considerazione sia la qualità attesa sia quella percepita.

Di solito, i questionari fanno uso di scale di valutazione di tipo verbale, mentre l'analisi dei dati è condotta interpretando le scale come

“numeriche”. Questo passaggio può generare due tipi di problemi: l’attribuzione di proprietà “più ricche” alle scale di misura da parte di chi ne fa l’elaborazione e l’arbitrarietà della codifica numerica delle informazioni raccolte. Se, dunque, da una parte la numerizzazione semplifica l’elaborazione dei dati, dall’altra ne allontana il significato dalla logica di chi li ha forniti.

Il problema della misurazione della qualità di un servizio va affrontato per gradi, cercando di stabilire:

- ✓ gli attributi da valutare, cioè le caratteristiche peculiari che influenzano l’erogazione del servizio e la loro importanza relativa dal punto di vista del cliente;
- ✓ i sistemi di misura più adeguati per il rilievo delle variabili in gioco;¹³
- ✓ le precauzioni da adottare per monitorare con continuità l’erogazione.

E’ chiaro quindi che la conoscenza e l’utilizzo delle metodologie per l’analisi e la misurazione della CS hanno oramai acquisito un’importanza fondamentale per le aziende orientate ad una politica di qualità, permettendo il costante monitoraggio delle prestazioni offerte sul mercato di riferimento, il controllo delle azioni di miglioramento, nonché la concreta attuazione di una politica di prevenzione della “non qualità”.

L’obiettivo di definire uno “standard” di valutazione, indipendente dal particolare contesto di applicazione, ha determinato la nascita, in letteratura, di numerose e diversificate metodologie per la valutazione della qualità dei servizi. Alcune nascono come esplicitazione di modelli concettuali disegnati per comprendere il meccanismo di valutazione,¹⁴ altre prendono spunto da analisi e sperimentazioni empiriche condotte su campioni e settori merceologici diversificati.

¹³ **ZANELLA A., CERRI M.**, *La misura di Customer Satisfaction: qualche riflessione sulla scelta delle scale di punteggio*. In *Valutazione della qualità e Customer Satisfaction: il ruolo della statistica. - Aspetti oggettivi e soggettivi della Qualità*. Atti della Giornata di studio promossa dall’AICQ, Bologna, 24 settembre 1999, pp. 217-231.

¹⁴ **PARASURAMAN A., ZEITHALM V.A., BERRY L.L.**, *A conceptual model of service quality and its implications for future research*. Cit., pp. 41-50.

Una prima distinzione va operata tra sistemi di misurazione diretta e indiretta:¹⁵

✓ **Misurazione diretta.** Si riferiscono a tutte le tecniche che comportano il diretto coinvolgimento del cliente, al quale viene chiesto di esprimere un giudizio sul proprio livello di soddisfazione nei confronti dell'azienda fornitrice; se, da una parte, sono tutti accomunati dall'utilizzo di un questionario appositamente strutturato, dall'altra, possono differenziarsi in base ad alcuni aspetti metodologici (campo d'indagine, metodo di rilevazione, frequenza di rilevazione, campionamento, ecc.);

✓ **Misurazione indiretta.** Consentono di ottenere una misura del livello di soddisfazione del cliente tramite la raccolta e l'analisi di informazioni ottenute indirettamente, per esempio, attraverso la gestione dei reclami o le impressioni del personale di contatto.

È opportuno, inoltre, segnalare che nella letteratura sul soddisfacimento del cliente il discorso verte quasi esclusivamente sulla qualità come è valutata da singoli o piccoli consumatori di beni o servizi, mentre viene praticamente ignorato il grande cliente interessato a grandi forniture ripetute nel tempo, al riguardo si veda la discussione e la differenziazione date in Vedaldi.

I modelli di misurazione ed interpretazione della CS possono essere convenientemente suddivisi attraverso la seguente macro-classificazione:

- 1) Modelli diretti formativi o compositivi;
- 2) Modelli diretti esplicativi o decompositivi;
- 3) Modelli a reti Bayesiane;
- 4) Modelli Strutturali

2.2.1 Nei modelli diretti formativi o compositivi

In tali modelli, considera il metodo fondato sulla definizione diretta della misura di Customer Satisfaction globale, come media ponderata di

¹⁵ COEN G. ET AL, *Customer Satisfaction nell'azienda di trasporto pubblico locale di Campobasso*. "Qualità", anno XXIX, n. 8, settembre 1999, pp. 66-79.

indicatori di Customer Satisfaction Marginali, attinenti a dimensioni specifiche del prodotto servizio (si considerino in particolare il SERVQUAL, SERVPERF).

Tali modelli a fronte di una forte semplicità applicativa ed interpretativa presentano l'inconveniente di non dare un'autonomia sperimentale al soddisfacimento o alla valutazione complessiva, di fatto quest'ultima viene definito attraverso una sintesi, al quanto arbitraria, delle valutazioni attinenti le varie dimensioni.

Tra i sistemi di misurazione diretta, uno dei più accreditati è il **Metodo SERVQUAL**, messo a punto da PARASURAMAN, ZEITHAML e BARRY.¹⁶

Nel modello alla base di questo metodo, la misurazione della soddisfazione del cliente si ottiene valutando la discrepanza tra percezioni ed aspettative; ciò risulta particolarmente efficace se si considera che la misurazione che si vuole effettuare riguarda una grandezza soggettiva (la soddisfazione), e non oggettiva, per la quale non esiste un sistema di riferimento uguale per tutti gli intervistati.

La base conoscitiva del metodo di PARASURAMAN, ZEITHAML e BARRY verte su uno studio multisettoriale a largo raggio che, condotto dagli autori alla fine degli anni '80, ha evidenziato quali fattori orientassero e determinassero nei clienti le valutazioni sulla qualità dei servizi.

Tale studio, è stato articolato essenzialmente in due indagini esplorative:¹⁷

- la prima è consistita in interviste ai dirigenti di quattro aziende note a livello nazionale, rappresentative di quattro differenti categorie di servizi: servizi bancari per il settore pubblico, carte di credito, intermediazione di titoli finanziari, riparazione e manutenzione di beni durevoli;
- la seconda è consistita in interviste a dodici *focus group* di clienti, tre per ognuna delle quattro categorie analizzate nell'indagine precedente, che fossero attuali o recenti utilizzatori dei servizi in esame.

¹⁶ PARASURAMAN A., ZEITHALM V.A., BERRY L.L., *A conceptual model of service quality and its implications for future research*. Cit., pp. 41-50.

¹⁷ PARASURAMAN A., *Servqual: applicazioni, messa a punto sul campo e risultati*. "De Qualitate", anno V, n. 6, giugno 1996, pp. 5-7.

I risultati della prima indagine esplorativa hanno condotto alla definizione di un modello concettuale, noto come “*Modello dei Gap*”,¹⁸ basato sulla convinzione che la qualità del servizio, così come percepita dai clienti, può essere definita come il grado di discrepanza (*gap*) tra le aspettative ed i desideri dei clienti e le loro percezioni.¹⁹

Tale divario o *gap* esterno tra qualità attesa e percepita (*gap 5*), misura il grado di soddisfazione del cliente nei confronti del servizio erogato dall’azienda; esso può essere scomposto in altri quattro scostamenti interni all’azienda stessa:

- ✓ *gap 1*: differenza tra le aspettative dei clienti e le percezioni del management su tali aspettative. Esso è attribuibile soprattutto alla carenza ed alla inesattezza delle informazioni esterne (ottenute, ad esempio, da ricerche di mercato), e delle informazioni interne (relative, ad esempio, alle comunicazioni tra *front line* e management);
- ✓ *gap 2*: differenza tra le percezioni che i dirigenti hanno delle aspettative del consumatore e le specifiche di qualità del servizio. Tale scostamento può essere legato ad una grande varietà di fattori quali: la scarsa specializzazione del personale, le fluttuazioni della domanda, un impegno inadeguato da parte del management per la qualità, ecc.;
- ✓ *gap 3*: differenza tra il servizio effettivamente erogato e le specifiche di qualità prefissate;
- ✓ *gap 4*: differenza tra il servizio promesso, definito attraverso le comunicazioni esterne ai clienti, e quello effettivamente erogato.

I risultati della seconda indagine esplorativa hanno consentito, in primo luogo, l’individuazione dei fattori chiave che determinano le aspettative dei clienti, quali: le comunicazioni tramite passaparola che si trasmettono da un consumatore all’altro, le esperienze personali dei clienti, le esperienze passate nell’uso del servizio, le comunicazioni esterne da parte dei fornitori. In secondo luogo, hanno portato all’individuazione dei criteri

¹⁸ COSTABILE M., *La misurazione della Customer Satisfaction: nuove ipotesi sul paradigma conferma/disconferma*. “Micro & macro marketing”, anno V, n. 3, dicembre 1996, pp. 475-501.

¹⁹ LONDRILLO L., *Il sistema Servqual come strumento di valutazione sulla qualità interna all’azienda*. “De Qualitate”, anno VIII, n. 10, novembre 1999, pp. 39-49.

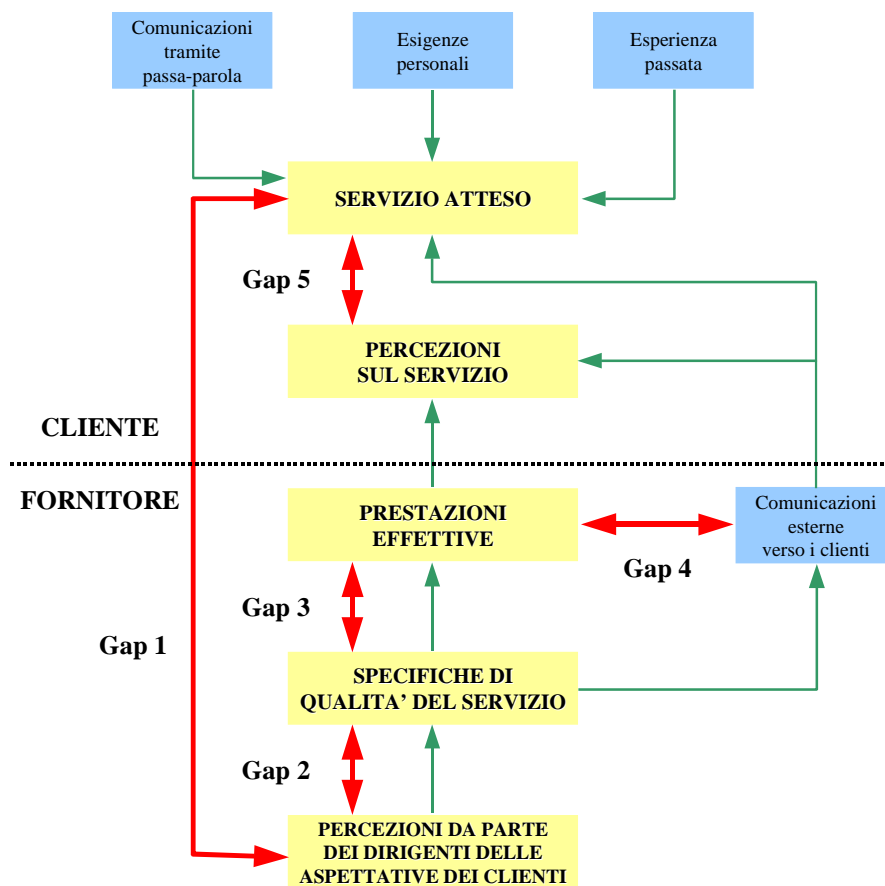
utilizzati dai clienti nel giudicare la qualità del servizio. Alla luce di queste considerazioni, le cinque componenti chiave della qualità risultano essere:²⁰

- ***aspetti tangibili***: aspetto delle strutture fisiche, dell'attrezzatura, del personale e degli strumenti di comunicazione;
- ***affidabilità***: capacità di presentare il servizio promesso in modo affidabile e preciso;
- ***capacità di risposta***: volontà di aiutare i clienti e di fornire prontamente il servizio;
- ***capacità di rassicurazione***: competenza e cortesia dei dipendenti e loro capacità di ispirare fiducia e sicurezza;
- ***empatia***: assistenza premurosa ed individualizzata che l'azienda presta ai singoli clienti.

Le indagini esplorative hanno condotto allo sviluppo del SERVQUAL quale strumento per la misurazione sistematica della qualità del servizio, ossia del **gap 5** (figura 2.1).

²⁰ Nella sua struttura originaria, il SERVQUAL prevedeva 10 dimensioni della qualità del servizio: Aspetti tangibili, Affidabilità, Capacità di risposta, Competenza, Cortesia, Credibilità, Sicurezza, Accesso, Comunicazione, Comprensione del cliente [ZEITHAML V.A., PARASURAMAN A., BERRY L.L., *Servire Qualità*. Milano, Mc Graw-Hill, 1991]. Successivamente, si è giunti alla formulazione a 5 dimensioni, condensandone alcune, in considerazione del contenuto delle voci affini [PARASURAMAN A., *La misura ed il controllo della qualità del servizio*. "De Qualitate", anno V, n. 5, maggio 1996, pp. 5-15].

Figura 2.1: “Modello concettuale del sistema *Servqual*”.



Fonte: Zeithaml V.A., Parasuraman A., Berry L.L., *Servire Qualità*. Milano, Mc Graw-Hill, 1991.

Il processo di misurazione si svolge somministrando un questionario suddiviso in tre sezioni che, valutando sia le aspettative (prima sezione) sia le percezioni (terza sezione) del cliente nei confronti del servizio ricevuto, consente di misurare la soddisfazione, anche attraverso i pesi assegnati all'importanza (seconda sezione) attribuita dall'intervistato alle diverse dimensioni.

Nella prima sezione, mediante 22 domande, si vogliono conoscere le aspettative dei clienti, intese non come semplice previsione del servizio e/o prodotto offerto dall'azienda in questione, ma come livello “quantitativo” di assolvimento dei vari aspetti del servizio e/o prodotto fornito da una generica azienda, al di sotto del quale si avrebbe insoddisfazione.

Nella terza, dedicata alle percezioni, si chiede di esprimere un giudizio sul prodotto e/o servizio offerto dall'azienda (per la quale viene fatta l'indagine), rispondendo a 22 domande, relative ai corrispondenti aspetti del servizio e/o prodotto analizzati nel caso delle aspettative.

Il questionario, infine, presenta un'altra sezione, la seconda, atta a verificare il giudizio dei clienti sull'importanza relativa delle cinque dimensioni della qualità del servizio erogato.

Il livello complessivo di soddisfazione dei clienti si ottiene mediante il calcolo del “*Punteggio Servqual Globale*”,²¹ un valore numerico che indica lo scostamento esistente tra la percezione che l'utente ha del servizio e le sue aspettative.

Esaminando i vari punteggi, un'azienda può non solo valutare la qualità globale del servizio com'è percepita dai clienti, ma individuare anche le dimensioni chiave e gli aspetti di tali dimensioni sui quali si dovrebbero concentrare gli sforzi di miglioramento delle percezioni.

I risultati ottenuti mediante il Metodo SERVQUAL possono essere utilizzati per diversi altri scopi: confrontare aspettative e percezioni dei clienti nel corso del tempo, confrontare i punteggi riportati dalla propria azienda con quelli della concorrenza, esaminare i segmenti della clientela con diverse percezioni della Qualità, valutare le percezioni sulla Qualità dei clienti interni.

Nel 1991, SCHVANEVELDT, ENKAWA e MIYIKAWA²² hanno elaborato il ***Two-Way Model***. L'intervistato valuta le numerose caratteristiche del servizio sotto due aspetti: uno “oggettivo” (con riferimento ad alcuni attributi della qualità) ed uno “soggettivo” (che implica la soddisfazione o l'insoddisfazione per il servizio ricevuto). Un questionario con coppie di domande relative ai due aspetti permette di classificare le risposte date dai clienti e valutare il servizio offerto.

²¹ **ORSINGHER C.**, *Strumenti di misurazione della Customer Satisfaction e della qualità nelle imprese di servizi: una rilettura critica*. “Sinergie”, anno XIV, n. 40, maggio/agosto 1996, pp. 209-225.

²² **SCHVANEVELDT S.J., ENKAWA T., MIYIKAWA M.**, *Consumer evaluation perspectives of service quality: evaluation factors and two-way model of quality*. “Total Quality Management”, vol. 2, n. 2, 1991.

Sulla falsariga delle ricerche condotte su SERVQUAL, CRONIN e TAYLOR (1992) hanno messo a punto un metodo denominato **SERVPERF**,²³ nell'intento di superare alcune delle difficoltà riscontrate nell'utilizzo del primo strumento. La principale novità di SERVPERF consiste nel richiedere all'intervistato solo valutazioni riguardanti le percezioni: secondo gli autori, questa procedura consente di ottenere migliori risultati di SERVQUAL, oltre a ridurre sensibilmente il numero delle domande poste.

TEAS, nel 1993, ha proposto il modello *Normed Quality* (NQ), al fine di precisare meglio il significato della componente delle aspettative.²⁴ Queste, infatti, possono essere interpretate dal cliente in vari modi: a livello ideale, prevedendo per ciascun attributo il livello massimo assoluto, oppure l'ideale realizzabile considerato relativamente alle particolari condizioni in cui il servizio può essere erogato.

Lo strumento del *Qualitometro*, proposto nel 1996 da FRANCESCHINI e ROSSETTO,²⁵ nasce con l'obiettivo di effettuare delle valutazioni ed un controllo "in linea" della qualità di un servizio, nel senso di un monitoraggio nel tempo dell'andamento del differenziale tra Qualità attesa e percepita δQ .²⁶

Tra gli aspetti interessanti di questa metodologia vi è quello di poter eseguire separatamente le "misure" della Qualità attesa Q_a e di quella percepita Q_p senza pericolo di inquinamenti reciproci. La prima è rilevata *ex-ante* la fruizione del servizio, la seconda, sullo stesso questionario, *ex-post*, al contrario di tutte le altre metodologie finora proposte, che richiedono una valutazione contemporanea *ex-post* sia delle attese che delle percezioni.

²³ CRONIN J.J., TAYLOR S.A., *SERVPERF versus SERVQUAL: reconciling performance-based and perceptions-minus-expectations measurement of service quality*. "Journal of Marketing", vol. 58, gennaio 1994.

²⁴ TEAS R.K., *Expectations, performance, evaluation and consumers perceptions of Quality*. "Journal of Marketing", vol. 57, luglio 1993.

²⁵ FRANCESCHINI F., ROSSETTO S., *Qualità nei servizi: un metodo per la valutazione e il controllo in linea del differenziale tra Qualità attesa e percepita*. "De Qualitate", anno V, n. 3, marzo 1996, pp. 53-64.

²⁶ FRANCESCHINI F., ROSSETTO S., *La valutazione e il controllo in linea della qualità dei servizi*. "De Qualitate", anno VI, n. 1, gennaio 1997, pp. 43-57.

Le dimensioni della qualità del servizio utilizzate nel Qualitometro sono quelle individuate nel modello SERVQUAL.

Il monitoraggio in linea è sviluppato mediante una carta per attributi di tipo **p** che consente di indagare su eventuali derive del differenziale δQ verso situazioni di “fuori controllo”.²⁷

La tabella 2.2 riporta in maniera sintetica le caratteristiche peculiari degli strumenti analizzati, mostrando i punti di affinità e le differenze sostanziali che li contraddistinguono.

Una differenza di rilievo riguarda il numero di enunciati che vengono sottoposti agli intervistati. Si passa da un minimo di 8 + 8 domande (8 per la valutazione delle attese e 8 per le percezioni) per il Qualitometro, ad un massimo di 10 + 10 + 10 + 10 + 10 domande richieste dal metodo NQ.

Determinare il numero di enunciati da inserire in un questionario è estremamente delicato, perché, se è vero che un numero maggiore di enunciati consente di disporre di più informazioni, è altrettanto vero che un numero troppo eccessivo stimola gli effetti di idiosincrasia e di stanchezza nella compilazione da parte degli intervistati, abbassando il livello di coinvolgimento e, quindi, l’attendibilità dell’informazione rilasciata.

²⁷ **FRANCESCHINI F., STANGALINI M.**, *Un’applicazione del metodo “Qualitometro” per la valutazione della qualità*. “De Qualitate”, anno IX, n. 4, aprile 2000, pp. 55-65; **FARNUM N.R.**, *Modern statistical quality control and improvement*. Belmont, Duxbury Press, 1997.

Figura 2.2: confronto di alcuni metodi per la valutazione qualità del servizio (Ripreso da Franceschini e Rossetto, 97)

Metodologie	SERVQUAL Revised Parasuraman Zeithaml Berry (1991)	Two-Way Schvaneveldt Enkawa Miyakawa (1991)	SERVPERF Cronin Taylor (1992)	Normed Quality Teas (1993)	Qualitometro Franceschini Rossetto (1996)
Fondamento teorico alla base del metodo	Teoria dei Gap: differenza tra percezioni ed attese mediata dall'effetto dei pesi dall'importanza assegnata ad ogni dimensione	I fattori latenti di valutazione sono suddivisi in: "oggettivi" (attributi della qualità) e "soggettivi" (livelli di soddisfazione)	Si considerano le sole "percezioni" senza la componente delle attese e senza i pesi dell'importanza	Il problema delle attese si distingue in attesa ideale I ed attesa realizzabile A	Si misurano le attese e le Percezioni in momenti separati. Tecniche MCDA.
Raccolta dati Caratteristiche del campione	Clienti di compagnie telefoniche, due assicurazioni, due banche	Banche, ristoranti, lavanderie, Supermercati	Due banche, due aziende di controllo antinquinamento, due lavanderie, due fast food	Tre grandi catene di supermercati	Utenti del servizio biblioteca del Dipartimento DISPEA
Numero di enunciati (attese più percezioni)	22+22	Non esplicitato	22	10+10+10+10+10	8+8
Scale per le risposte	Semantica differenziale a 7 punti	Semantica a 5 punti	Semantica differenziale a 7 punti	Semantica differenziale a 7 punti	Comparativa a 7 punti
Importanza delle dimensioni	Valutazione dei pesi a somma costante	Non richiesto	Valutazione dei pesi a somma costante	Valutazione dei pesi a somma costante	Comparativa a 7 punti
Modalità di consegna del questionario	Postale	Non precisato	Postale	Colloquio	Attese prima della fruizione del servizio e percezioni dopo la ricezione
Livello di intrusione nei confronti del misurando	Alto	Medio	Alto	Alto	Basso
Effetto di Idiosincrasia	Alto	Medio	Medio	Alto	Basso
Tipo di pre-elaborazione dei dati	Scalarizzazione	Scalarizzazione	Scalarizzazione	Scalarizzazione	Senza scalarizzazione
Analisi dei dati	Analisi fattoriale seguita con rotazione obliqua	Analisi fattoriale	Analisi fattoriale seguita con rotazione obliqua	Analisi fattoriale seguita con rotazione obliqua	Metodi di MCDA, carte di controllo di tipo "p"
Affidabilità (coefficiente di Cronbach)	Da 0.8 a 0.93	Non esplicitato	Da 0.63 a 0.98	Calcolati altri coefficienti di validità ed affidabilità	Calcolo della qualità "globale" come indicatore di affidabilità
Numero delle dimensioni	Cinque: Aspetto, affidabilità, capacità di risposta, capacità di rassicurazione, empatia	Cinque: Prestazione, sicurezza, completezza, facilità di utilizzo, emotività/ambiente	Cinque: Aspetto, affidabilità, capacità di risposta, capacità di rassicurazione, empatia	Cinque: Aspetto, affidabilità, capacità di risposta, capacità di rassicurazione, empatia	Cinque: Aspetto, affidabilità, capacità di risposta, capacità di rassicurazione, empatia

2.2.2 Modelli Diretti Esplicativi o Decompositivi

In tali modelli, si considera il metodo, fondato sulla misurazione diretta della CS globale tramite una scala di punteggi convenzionale, che formula un'ipotesi sul sottostante costrutto concettuale attraverso il collegamento, mediante un modello di regressione, dei precedenti punteggi con indicatori relativi a dimensioni specifiche o marginali, sia del prodotto servizio che dell'unità statistica (soggetto). È il caso sia dei modelli di regressione applicati a variabili indicatrici o loro trasformazioni secondo metodologie di scaling quali quella di Thurstone, sia dei modelli di regressione logistica in cui la risposta è la probabilità di soddisfacimento globale, condizionata da specifici valori delle variabili descrittive del costrutto sottostante, sia, infine, dei modelli riconducibili alla Rasch Analysis.

Un'assunzione teorica che consente, fra l'altro, di ovviare al problema della convenzionalità della scala scelta per raccogliere i giudizi, suggerisce di considerare la CS nei termini di variabile (o tratto) latente. In altri termini l'informazione tratta dalle osservazioni (variabile manifesta) non è immediatamente riferibile alla variabile oggetto di indagine (la CS in questo caso), che viceversa è inosservabile; tale informazione viene pertanto utilizzata mediandola con altre assunzioni teoriche.

L'ipotesi è che i fattori che influenzano la propensione del soggetto per una determinata categoria di risposta piuttosto che un'altra, nel giudizio sul grado di soddisfacimento per ogni prefissato attributo, siano due: un fattore specifico dell'attributo (∂) ed un fattore specifico del soggetto (β). Quest'ultimo fattore in particolare rende ragione almeno in parte delle diversità fra soggetti. Le ipotesi di base che soggiacciono al modello di Rasch sono la unidimensionalità e la separabilità dei parametri.

L'unidimensionalità presuppone l'esistenza di una medesima metrica a cui possono essere ricondotti i fattori precedentemente introdotti. La separabilità consente di procedere efficacemente alla stima dei parametri degli attributi senza che vengano coinvolti i parametri dei soggetti, e viceversa. In altri termini si suppone che tutta l'informazione campionaria

per la stima del parametro (β) è contenuta nei punteggi degli attributi , e viceversa per (∂)

È importante osservare che, nella prospettiva delle determinazioni di un efficiente indicatore standardizzato della CS, la selezione stessa degli attributi può e deve essere finalizzata alla pertinenza degli stessi a quest'obiettivo; pertanto gli attributi dovranno essere omogenei, misurando tutti lo stesso tratto latente, ma nel contempo eterogenei, essendo differenziati in modo da non aver altro in comune che il tratto latente.

Se indichiamo con (β) l'ideale espressione numerica del fattore attributo e con (∂) l'espressione numerica del fattore beta, l'ipotesi semplice che si effettua con il modello di Rasch e che i due fattori interagiscono tra loro secondo la formula additiva seguente $\beta + \partial$.

Nel caso più semplice di giudizio di tipo dicotomico espresso dall'intervistato (soddisfatto – non soddisfatto) si può pensare di trasformare la formula precedentemente introdotta in un numero che rappresenti la probabilità che l'intervistato in questione risponda con un giudizio positivo alla domanda postagli.

In particolare utilizzando la trasformazione logistica si ottiene $\exp(\beta + \partial)/[1+\exp(\beta + \partial)]$.

È questo il modello di Rasch per variabili dicotomiche.

Nel caso più complicato in cui le domande prevedono più di 2 categorie di risposta ordinate, il modello è detto politomico. In questo caso l'ipotesi che si postula per l'interazione tra i fattori del modello è la seguente formula bilineare: ($\eta_h\beta + v_h\partial + \gamma\beta\partial + \delta_h$).

Ipotizzando l'assenza di interazione tra soggetto ed attributo e la seguente serie di uguaglianze $\eta_h = v_h = \emptyset_h$ si ottiene la seguente versione politomica del modello di Rasch per la probabilità di scelta della generica categoria di risposta c_h :

$$[\text{Exp}(\emptyset_h(\beta + \partial) + \delta_h) / \sum_{h=0}^m \text{exp}(\emptyset_h(\beta + \partial) + \delta_h)]$$

2.2.3 Modelli Di Reti Bayesiane

Si mostra come tali modelli, basati su grafi orientati aciclici e riferiti a variabili manifeste, tra le quali una attinente alla CS globale, tipicamente misurata su una scala di punteggi convenzionale, possano servire al fine di stabilire per quest'ultima un costrutto concettuale di casualità. Ciò è ottenuto dall'esame sistemico della verosimiglianza dei dati osservati nei nodi della rete, ciascuna condizionatamente alla configurazione di uno o più nodi antecedenti. Il metodo sembra particolarmente adatto al caso di variabili manifeste di tipo qualitativo.

2.2.4 Modelli Strutturali

Avviare un programma di qualità, come è uno studio di CS, richiede l'uso di strumenti adeguati; molte metodologie si limitano ad utilizzare medie aritmetiche o al massimo analisi di correlazione, ma l'uso di strumenti di calcolo qualitativamente poveri può portare a risultati non utilizzabili, cioè non operativi, per l'impossibilità di collegare gli indici di CS agli indicatori economici.

Questi metodi, utilizzati abitualmente nelle ricerche di mercato, sono ormai superati da tecniche più avanzate e sofisticate: il metodo delle equazioni strutturali, il metodo dei minimi quadrati parziali (PLS). Tali strumenti consentono di risolvere le difficoltà intrinseche nell'elaborazione dei dati di CS:

- ✓ la distribuzione dei dati, non gaussiana ma asimmetrica;
- ✓ l'elevato numero di variabili da gestire;
- ✓ l'alta collinearità tra le componenti della qualità, che impedisce, con i metodi tradizionali, di isolare gli effetti delle singole variabili sulla CS.

Inoltre, essi consentono una precisione di calcolo superiore o, a parità di precisione, di ridurre il numero di casi necessari (che si traduce in una riduzione dei costi di rilevazione).

Dagli studi di CS dipendono alcune delle scelte strategiche che l'azienda deve effettuare: da qui l'importanza di una metodologia efficace e collaudata, in grado di fornire al management le informazioni di cui ha bisogno per effettuare le scelte migliori.

I *Modelli ad Equazioni Strutturali* forniscono ai ricercatori un'utile metodologia per valutare le principali componenti della verifica di una teoria, quali l'adeguatezza della misurazione delle variabili e l'analisi delle relazioni ipotizzate tra di esse; superano la logica "descrittiva" nella valutazione della qualità, in favore di una logica "esplicativa", che tiene conto dei *nessi causali* tra le dimensioni della qualità.²⁸

Gli indicatori trasformati in variabili sono di fatto legati da relazioni di natura logica; la spiegazione dipende dall'interpretazione di queste ultime per l'attribuzione di un "senso".

L'analisi causale è particolarmente idonea ad evidenziare i nessi tra le concrete manifestazioni empiriche ed i fenomeni rilevati tramite gli indicatori di qualità.

Poiché nella maggior parte dei casi il processo causale non può limitarsi ad una singola relazione, ma molte cause sono collegate fra loro e con vari effetti, lo strumento più idoneo ad effettuare queste valutazioni è rappresentato dai modelli ad equazioni strutturali.

L'impiego di tali modelli negli studi di marketing è oggi riconosciuto a livello internazionale, perché essi costituiscono un approccio generale per integrare la fase di costruzione di una teoria con gli stadi di ipotesi e di verifica empirica;²⁹ come è stato evidenziato nel paragrafo precedente,

²⁸ **OLIVERI A.**, *La valutazione della qualità dei servizi sanitari: dalla descrizione alla spiegazione*. In Atti del I Convegno Internazionale: *Gestione della Qualità Totale nelle strutture sanitarie: dalla teoria alla pratica*. Troina (EN), 17 aprile 1998, pp. 1-14.

²⁹ **TASSINARI G. ET AL.**, *La soddisfazione del cliente dei servizi di segreteria universitaria: un modello ad equazioni strutturali*. In *Valutazione della qualità e Customer Satisfaction: il ruolo della statistica. - Aspetti oggettivi e soggettivi della Qualità*. Atti della Giornata di studio promossa dall'AICQ, Bologna, 24 settembre 1999, pp. 291-316.

attualmente, rappresentano lo strumento più efficace per la costruzione dei maggiori indicatori nazionali di CS.

I modelli ad equazioni strutturali possono essere utilizzati come sofisticati strumenti di misura per generare delle scale e degli indici di cui poi valutare l'affidabilità, come procedure per valutare le varie forme di validità delle misure, o come metodologie per compiere test di ipotesi, predire l'andamento di variabili, o supportare inferenze su cause ed effetti.³⁰

Non mancano in effetti critiche a tale metodo collegate alla mancanza di consistenza ed efficienza asintotica degli stimatori dei parametri ed alla constatazione che gli stimatori delle variabili latenti e degli errori non rispettano, in generale, le condizioni di non correlazione assunte tipicamente per i modelli lineari strutturali a variabili latenti.

Tali modelli tengono conto del carattere autonomo ma relazionale del soddisfacimento complessivo e tendono a stabilire direttamente, attraverso le variabili manifeste, il costrutto concettuale in cui è incapsulata la nozione di CS.

A rigore sono giustificati soprattutto come modelli confermativi di un costrutto già individuato.

³⁰ Ibidem.

CAPITOLO TERZO

EQUAZIONI STRUTTURALI: FORMULAZIONE DEL MODELLO

Sommario: - **3.1** Premessa - **3.2** Tipi di relazioni causali tra le variabili - **3.2.1** Relazione diretta - **3.2.2** Relazione reciproca - **3.2.3** Relazione spuria - **3.2.4** Relazione indiretta - **3.2.5** Relazione condizionata (interazione) - **3.3** «*Lisrel*» ed i modelli di equazioni strutturali - **3.4** La logica e le fasi di *Lisrel* - **3.5** La formulazione del modello secondo la notazione *Lisrel* - **3.6** La rappresentazione grafica del modello - **3.7** Il modello strutturale - **3.8** Il modello di misurazione per le variabili endogene - **3.9** Il modello di misurazione per le variabili esogene - **3.10** Le restrizioni del modello.

Premessa.

Molti fenomeni economici, più o meno complessi, non sono spiegabili con modelli ad una sola equazione, infatti, spesso, allo scopo di avere un'accurata rappresentazione del fenomeno, è necessario ricorrere ad un sistema di equazioni.

*Se si pensa che un sistema economico è formato da più parti tra loro interconnesse, allora è chiaro che non sarebbe possibile studiare una singola parte separatamente dal resto, quindi lo studio andrebbe riservato solo all'intero complesso di parti.*³¹

In termini formali, ciò equivale a lavorare con modelli di più equazioni in qualche modo collegate, o per la presenza delle stesse variabili nelle diverse relazioni e/o da legami tra le componenti stocastiche delle varie relazioni.

³¹ PERSICO P., VINCI S., *Principi di Econometria*. Napoli, Liguori, 1981.

I *Modelli*³² *ad Equazioni Strutturali*, nell'approccio comunemente noto con il nome di “*Lisrel*” (*Linear Structural Relationship*), rappresentano uno dei più efficaci strumenti di analisi multivariata attualmente disponibili in letteratura nel campo delle *variabili metriche*.³³

Essi assumono la forma di sistemi di equazioni algebriche, dove ogni variabile dipendente è espressa in funzione delle variabili indipendenti su di essa agenti.

Questi modelli sono stati elaborati agli inizi degli anni settanta e rappresentano la sistemazione logica, prima ancora che statistica, di tecniche di analisi multivariata, riconducendo ad un unico modello approcci scientifici finora distinti e non comunicanti quali, l'analisi fattoriale, i modelli causali ed i modelli di misurazione. In particolare, rappresentano la più completa risposta al problema di operativizzare, in termini di ricerca e di verifica empirica, la nozione di causalità. Si tratta dunque della generalizzazione di quelli che, negli anni sessanta, venivano definiti *modelli causali*.

Grazie alla possibilità di poter includere nel modello teorico e nella trattazione statistica anche delle *variabili latenti*,³⁴ i modelli di equazioni

³² Intendiamo per “modello”: “... *l'espressione semplificata e formalizzata del processo causale che si pensa esista nella realtà, dove riduzione della complessità della teoria e formulazione secondo un insieme sintattico di simboli ne rappresentano i due elementi qualificanti*”. [CORBETTA P., *Metodi di analisi multivariata per le scienze sociali*. Bologna, Il Mulino, 1992].

³³ Classificando le variabili secondo il loro livello di misurazione, esse vengono distinte in: *nominali*, quando presentano stati discreti non ordinabili; *ordinali*, quando presentano stati ordinabili ma non si dispone di una unità di misura, per cui non è possibile rilevare la distanza esistente fra tali stati; *ad intervalli*, quando esiste una unità di misura, ma non esiste uno zero assoluto; *cardinali*, quando esiste anche uno zero assoluto. Le ultime due tipologie di variabili descritte costituiscono il gruppo delle “*variabili metriche*”, caratterizzate dall'esistenza di una unità di misura (o di conto). [STEVENS S.S., *On the theory of scales of measurement*. “Science”, n. 103, 1946, pp. 670-680].

³⁴ Da un punto di vista concettuale, occorre precisare la distinzione che intercorre fra variabili *latenti* e variabili *osservate*. Si definiscono *variabili latenti*: “*dei costrutti teorici che non sono direttamente osservati, ma che hanno implicazioni per le relazioni fra le variabili osservate*” [GOLDBERGER A.S., DUNCAN O.D., *Structural Equation Models in the Social Science*. New York, Seminar Press, 1973]. Per la loro stessa natura, le variabili latenti non sono osservabili e, quindi, non “direttamente” misurabili; status socio-economico, pregiudizio razziale, intelligenza, aspettative economiche, *customer satisfaction*, sono solo alcuni esempi di concetti teorici per i quali non esistono strumenti di misura. Sono tuttavia misurabili delle variabili specifiche legate al più generale concetto teorico sottostante (e da questo causalmente prodotte); ad esempio, dall'analisi del comportamento del cliente nei confronti dell'offerta di servizi da parte di una determinata organizzazione è possibile far discendere una misura della sua soddisfazione e, quindi, un complesso di decisioni e di comportamenti che porteranno all'offerta di un servizio coerente [TASSINARI G. ET AL, *La soddisfazione del cliente dei servizi di segreteria universitaria*:

strutturali forniscono una nuova formulazione alla vasta famiglia delle tecniche di *analisi fattoriale*.³⁵

3.2 Tipi di relazioni causali tra le variabili.

Prima di definire le differenti relazioni di causalità fra le variabili, è utile chiarire la distinzione fra covariazione e causazione.

Si parla di **covariazione** (covarianza, correlazione o associazione), quando si osserva che due variabili presentano variazioni concomitanti: al variare dell'una varia anche l'altra, senza che ciò implichi un nesso logico tra di esse.

Si parla invece di **causazione** quando è implicata la nozione di "produzione": *"Se X è una causa di Y, è possibile che una trasformazione in X produca una trasformazione in Y, e non semplicemente che una trasformazione in X sia seguita da, o associata a, una trasformazione in Y"* [Blalock 1961].

Dalla definizione si evince che gli elementi in più presenti nel concetto di causazione rispetto a quello di covariazione sono due. Da una parte, la *direzionalità* (o asimmetria) dell'azione: nella causazione esiste una causa ed un effetto, il variare di una variabile precede il variare dell'altra, mentre nella covariazione esiste solo la concomitanza della variazione. Dall'altra, quello di *legame diretto* tra le due variabili, nel senso che il variare di una variabile è dovuto - e non esclusivamente associato - al variare dell'altra.

E' possibile, dunque, che esista covariazione senza che esista causazione.

Il meccanismo causale fra due variabili può essere essenzialmente di cinque tipi (per una loro rappresentazione grafica si veda la figura 3.1):

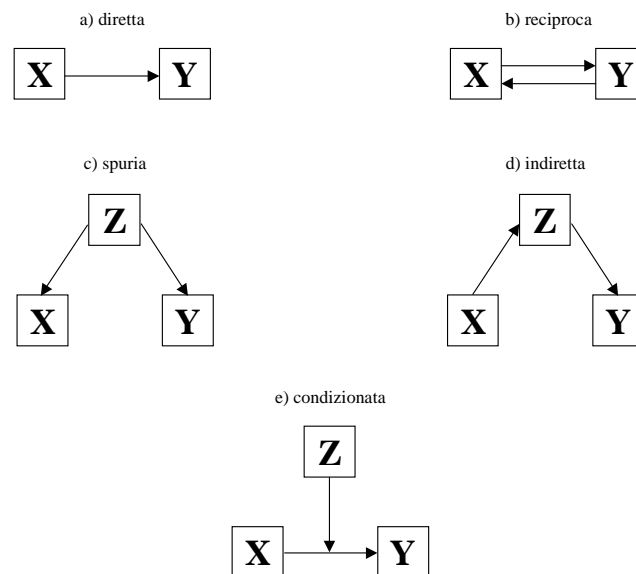
- Relazione diretta;

un modello ad equazioni strutturali. In Valutazione della qualità e Customer Satisfaction: il ruolo della statistica. - Aspetti oggettivi e soggettivi della Qualità. Atti della Giornata di studio promossa dall'AICQ, Bologna 24 settembre 1999, pp. 291-316].

³⁵ L'**analisi fattoriale** rappresenta il tentativo di scoprire se le correlazioni esistenti fra un certo numero di *variabili osservate* possano essere spiegate da un numero inferiore di *variabili latenti* o "fattori". [CORBETTA P., *Metodi di analisi multivariata per le scienze sociali*. Cit.].

- Relazione reciproca;
- Relazione spuria;
- Relazione indiretta;

Figura 3.1: “Rappresentazione grafica dei cinque possibili tipi di relazione causale fra due variabili X e Y”.



Fonte: Corbetta P., *Metodi di analisi multivariata per le scienze sociali*. Bologna, Il Mulino, 1992.

- Relazione condizionata.

3.2.1 Relazione diretta.

Si tratta del tipo di relazione più immediato: due variabili sono legate da una relazione causale *diretta* quando un mutamento nella variabile “causa” produce un mutamento nella variabile “effetto”.

Come già accennato in precedenza, gli elementi essenziali di questo tipo di relazione sono:

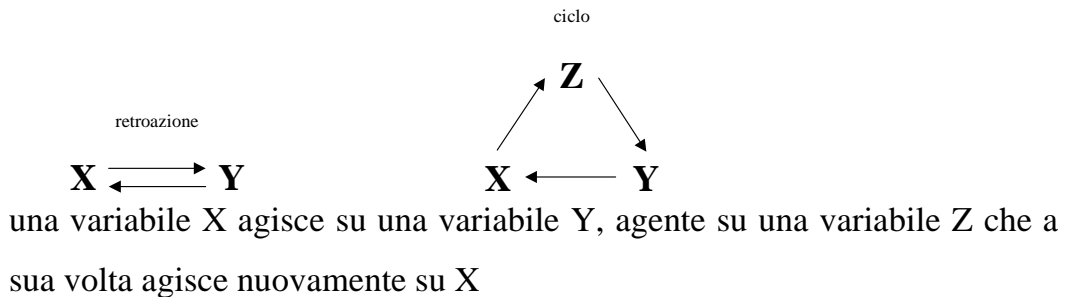
- l’asimmetria o la direzionalità del rapporto, senza la quale avremmo una relazione “reciproca” (*infra* par. 3.2.2);
- l’inerenza del concetto di produzione, senza il quale avremmo solo covariazione, e cioè una relazione “spuria” (*infra* par. 3.2.3);

- c) l'immediatezza del nesso, senza la quale la relazione causale sarebbe mediata da un'altra variabile, sarebbe cioè una relazione "indiretta" (*infra* par. 3.2.4).

3.2.2 Relazione reciproca.

Si parla di *causazione reciproca* o *retroazione*, quando non sussiste una distinzione fra variabile causa e variabile effetto, ma le variabili si influenzano reciprocamente; dunque, viene meno l'asimmetria del rapporto.

Accanto alla situazione di retroazione abbiamo quella di *ciclo*, dove



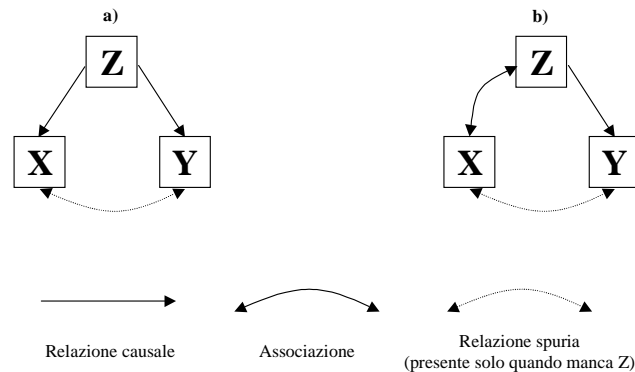
3.2.3 Relazione spuria.

Si definisce relazione *spuria* quella che sussiste tra variabili che "presentano covariazione pur in assenza di causazione".

Il meccanismo causale sottostante una relazione spuria è il seguente: la covariazione fra le due variabili considerate X e Y è provocata da una terza variabile Z che agisce causalmente sia su X che su Y.

Come si può osservare dalla figura 3.2a), in tutti i casi di relazione

Figura 3.2: “Rappresentazione grafica di relazioni spurie tra le variabili X e Y”.



Fonte: Corbetta P., *Metodi di analisi multivariata per le scienze sociali*. Bologna, Il Mulino, 1992.

spuria tra X e Y, la variabile Z (chiamata *variabile di controllo*) è realmente connessa causalmente con la X e con la Y, mentre queste due variabili covariano a causa della loro relazione con la Z, ma non hanno legame causale fra loro. La Z, dunque, rappresenta la “causa comune” alle spalle delle due variabili covarianti.

La covariazione tra X e Y dipende chiaramente da Z, infatti, se la variazione di Z sparisce, sparisce anche la loro relazione. Supponiamo, ad esempio, che Z sia legata positivamente con X e positivamente con Y. Elevati valori di Z produrranno elevati valori di X ed anche elevati valori di Y. Simmetricamente per i valori bassi di Z. Di conseguenza, quando è presente l’azione di Z, le unità di analisi aventi alti valori di X presenteranno anche alti valori di Y, e simmetricamente per valori bassi, per cui le due variabili X e Y covarieranno.

In conclusione, è possibile affermare che la covariazione spuria tra X e Y emerge solo se *non* viene collocata esplicitamente nel modello teorico la variabile Z. Se invece si è consapevoli della sua influenza e la si tiene sotto controllo, l’effetto spurio sparisce: nel primo caso perché, venendo a mancare l’effetto di Z su X e su Y, viene meno anche quello spurio fra X e Y; nel secondo caso, in quanto l’effetto spurio fra X e Y è sostituito da quelli reali fra Z e X e fra Z e Y.

Una situazione di relazione spuria si può verificare anche nel caso in cui il legame fra la Z e la variabile indipendente X è di semplice covariazione e non di causalità. Il meccanismo è rappresentato graficamente in figura 3.2b).

Questa particolare forma di relazione spuria fra X e Y viene definita *relazione congiunta (joint)*, perché la causazione sulla Y non proviene dalla X, ma da un'altra variabile a questa correlata.

3.2.4 Relazione indiretta.

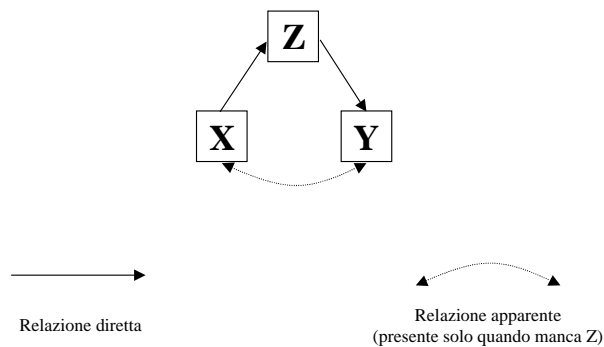
Una relazione causale *indiretta* tra due variabili si presenta quando il loro legame causale è mediato da una terza variabile Z.

Come nel caso di relazione spuria, anche in presenza di relazione causale indiretta la covariazione fra X e Y è in realtà determinata da una terza variabile Z, ma, in questo caso, Z funge da “ponte” fra X e Y, che attraverso di essa risultano causalmente connesse, mentre nel caso precedente, la relazione fra X e Y è inesistente.

Il meccanismo causale è descritto in figura 3.3: X agisce su Z, la quale agisce su Y.

La variabile Z, dunque, interviene nel rapporto tra X e Y e per questo viene definita *variabile interveniente*.

Figura 3.3: “Rappresentazione grafica di relazione indiretta tra le variabili X e Y”.



Fonte: Corbetta P., *Metodi di analisi multivariata per le scienze sociali*. Bologna, Il Mulino, 1992.

3.2.5 Relazione condizionata (interazione).

Il caso della relazione *condizionata* è quello in cui la relazione fra due variabili cambia a seconda del valore assunto da una terza variabile ed è rappresentato graficamente dall'ultimo caso in figura 3.1.

L'azione della variabile Z si esercita non su X o su Y, ma sul legame che lega X a Y; è per tale ragione che, in generale, è possibile affermare che ci troviamo in presenza di un fenomeno di *interazione* fra le variabili implicate.

3.3 «Lisrel» ed i modelli di equazioni strutturali.

Il termine *Lisrel* nasce come nome di un *software* messo a punto dallo statistico-psicometrico svedese Karl Jöreskog nei primi anni settanta per stimare, con il metodo della massima verosimiglianza, i coefficienti strutturali dell'analisi fattoriale.

La sua applicazione ha superato rapidamente i confini dell'analisi fattoriale, diventando una procedura generale per i modelli basati su sistemi di equazioni strutturali.

L'approccio *Lisrel*, mutuando dalla tradizione scientifica psicometrica la nozione di variabile latente e dall'econometria la nozione di rete di relazioni causali tra le variabili, è costituito da due parti:

- ✓ **Modello di misurazione**: consente di spiegare come le variabili latenti vengono misurate tramite le variabili osservate e serve per determinare la validità e l'attendibilità di tale misurazione. Questo modello risponde all'esigenza di risolvere il problema della *misurazione* di variabili che rappresentano concetti teorici non osservabili direttamente, o per le quali non esistono adeguati strumenti di misura.
- ✓ **Modello strutturale**: specifica le relazioni causali tra le variabili latenti e serve per determinare gli effetti causali e l'ammontare della varianza non spiegata. Si cerca, con questo modello, di rispondere alla necessità di disporre di strumenti e metodi per saggiare empiricamente l'esistenza dei nessi causali fra le variabili, formulati in sede teorica.

Per *modello di equazioni strutturali* si intende: “un modello stocastico nel quale ogni equazione rappresenta un legame causale, piuttosto che una mera associazione empirica” [Goldberger 1972].

L'unità costitutiva di un modello di equazioni strutturali è *l'equazione di regressione*, ma, in questo contesto, a livello teorico, viene data ad essa un'interpretazione di carattere causale.

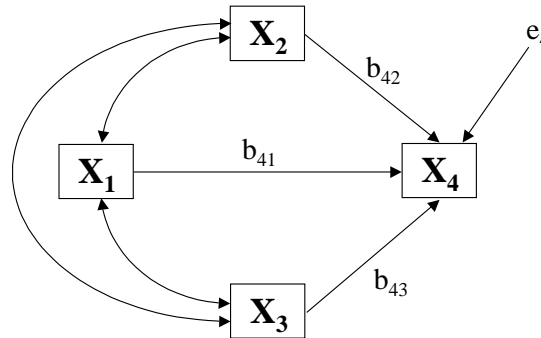
Se, ad esempio, esiste una covariazione fra due variabili spurie,³⁶ l'equazione di regressione del tipo $Y = a + bX$ ha un senso statistico, in quanto esprime un nesso realmente esistente; ha anche un senso agli effetti della predizione della Y , in quanto, data l'esistenza della relazione e data la conoscenza dei suoi parametri (cioè a e b) su un certo insieme di unità di analisi, è possibile stimare la Y , nota la X .

Tuttavia, ciò non ha nulla a che fare con il processo di *spiegazione causale*; l'equazione di regressione, dunque, se è utile ai fini della “predizione” della variabile dipendente Y , è irrilevante ai fini della “spiegazione” della stessa variabile.

³⁶ Cfr. paragrafo 3.2.3.

Nei modelli di equazioni strutturali siamo interessati all'equazione di

Figura 3.4: “Rappresentazione grafica del modello di regressione”.



Fonte: Corbetta P., *Metodi di analisi multivariata per le scienze sociali*. Bologna, Il Mulino, 1992.

regressione solo nella misura in cui è possibile attribuirle un significato di nesso causale. Essa viene definita *equazione strutturale* ed esprime, attraverso la formulazione matematica, la relazione esistente fra una variabile dipendente e diverse variabili indipendenti.

Consideriamo la seguente equazione strutturale:

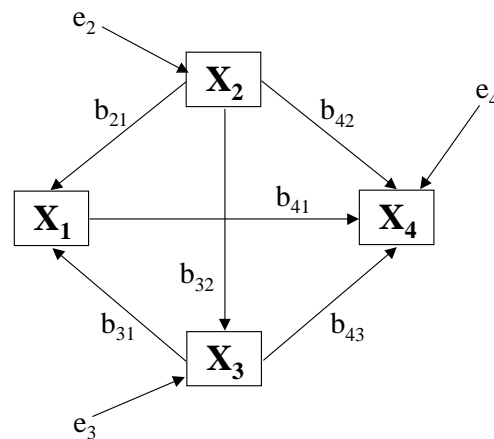
$$X_4 = b_{41}X_1 + b_{42}X_2 + b_{43}X_3 + e_4$$

La variabile X_4 è messa in relazione con le variabili X_1 , X_2 e X_3 .

In questa formulazione, come in tutte le successive relative ai modelli di equazioni strutturali, le variabili vengono espresse in termini di *scarti dalle rispettive medie*, dunque, nell'equazione di regressione non compare l'intercetta a .

Nel momento in cui questa equazione rappresenta un processo causale, può essere rappresentata in forma grafica, come si evince dalla figura 3.4.

Figura 3.5: “Rappresentazione grafica del modello di equazioni strutturali”.



Fonte: Corbetta P., *Metodi di analisi multivariata per le scienze sociali*. Bologna, Il Mulino, 1992.

Ciascuna freccia unidirezionale rappresenta una relazione causale, mentre ogni freccia ad arco bidirezionale indica una semplice interazione. I coefficienti di regressione vengono definiti *parametri strutturali* e vengono collocati sulle frecce indicanti i nessi causali.

Il modello causale appena descritto, definisce il **modello di regressione**, in quanto la variabile dipendente è influenzata da un certo numero di variabili indipendenti, che fra loro non presentano relazioni di causalità, ma solo di associazione.

In realtà, si prendono in considerazione solo le relazioni causa-effetto agenti dalle variabili indipendenti sulla dipendente, senza prendere in considerazione i meccanismi causali che possono esistere anche fra le indipendenti.

Supponendo che la variabile X_3 sia influenzata da X_1 e da X_2 , e che a sua volta quest'ultima sia influenzata da X_1 , il modello complessivo, con le equazioni che esprimono anche le dipendenze di X_2 e di X_3 , diventa il seguente (figura 3.5):

$$X_2 = b_{21}X_1 + e_2$$

$$X_3 = b_{31}X_1 + b_{32}X_2 + e_3$$

$$X_4 = b_{41}X_1 + b_{42}X_2 + b_{43}X_3 + e_4$$

Il sistema di equazioni appena descritto, dove ogni equazione rappresenta un nesso causale, è quello che viene definito **modello di equazioni strutturali**. Si tratta di un insieme di nessi causali fra variabili, formalizzati nel loro complesso mediante un sistema di equazioni

algebriche. In una formulazione generale, il modello può essere rappresentato come segue:

$$X_1 = b_{12}X_2 + b_{13}X_3 + \dots + b_{1k}X_k + e_1$$

$$X_2 = b_{21}X_1 + b_{23}X_3 + \dots + b_{2k}X_k + e_2$$

.....

$$X_k = b_{k1}X_1 + b_{k2}X_2 + \dots + b_{k,k-1}X_{k-1} + e_k$$

Ciascuna equazione esprime il legame causale esistente fra una variabile dipendente, riportata nel primo membro dell'equazione, ed un certo numero di variabili indipendenti, presenti nel secondo membro. I valori dei coefficienti b ci indicano di *quanto* la variabile dipendente è influenzata da ciascuna variabile indipendente. Le equazioni saranno tante quante sono le variabili dipendenti.

Sia nel modello di regressione che nel modello di equazioni strutturali, il coefficiente b_{42} va interpretato come l'*effetto diretto* di X_2 su X_4 , vale a dire, come la variazione causata su X_4 dalla variazione di un'unità in X_2 , tenute costanti le altre variabili. Tuttavia, se X_2 agisce anche su X_3 , la quale agisce a sua volta su X_4 , una modifica di X_2 produrrà su X_4 due effetti: uno diretto, ma anche uno indiretto attraverso l'azione che passa attraverso X_3 .

Il modello di regressione non prende in considerazione le interazioni causali fra X_2 e X_3 e ciò impedisce di cogliere tale *effetto indiretto*, dunque, di conseguenza, attribuisce tutto l'effetto (*effetto totale*) di X_2 su X_4 al solo effetto diretto.

In conclusione, l'approccio a più equazioni è l'unico che fornisce una rappresentazione, per quanto semplificata, dei processi reali, perché tiene conto non solo della molteplicità delle cause che agiscono su una variabile dipendente (analisi multivariata), ma anche delle connessioni esistenti fra le diverse cause.

I processi reali vanno intesi, infatti, come una rete complessa di interazioni, e l'approccio a più equazioni permette appunto di definire la *struttura* di tale rete. Da qui la definizione di modelli di equazioni "strutturali". Simmetricamente, la singola equazione componente il

sistema, viene definita “equazione strutturale” ed i coefficienti b “parametri strutturali”.

Il passaggio dall’approccio per singole equazioni a quello per sistemi di equazioni, cioè dal “modello di regressione” ai “modelli di equazioni strutturali”, comporta una completa revisione del processo di stima dei parametri del modello. Mentre, infatti, nel caso del modello di regressione è possibile procedere con il metodo di stima dei minimi quadrati (OLS), nel caso di un modello costituito da più equazioni, dove le variabili indipendenti di un’equazione risultano le dipendenti di un’altra, si introducono enormi complicazioni nel processo di stima dei coefficienti b , in quanto una delle condizioni essenziali della stima dei minimi quadrati, quella della indipendenza fra gli errori e e le variabili indipendenti X , normalmente non si verifica più.³⁷

Il fatto che le variabili possano essere, nello stesso modello di equazioni strutturali, contemporaneamente dipendenti ed indipendenti, richiede una modifica della terminologia finora utilizzata, al fine di non generare confusione. Parleremo dunque di ***variabili esogene*** per indicare tutte le variabili “esterne” al modello strutturale, che in esso intervengono sempre e solo come variabili indipendenti. Verranno definite ***variabili endogene*** tutte le variabili “interne” al modello, che alternativamente, nelle varie equazioni, possono comparire come dipendenti o indipendenti.

Nella notazione *Lisrel* indicheremo con la lettera ***X*** le variabili esogene e con la ***Y*** quelle endogene.

Le variabili esogene vengono anche definite “predeterminate”, per sottolineare il fatto che il loro valore è determinato *al di fuori* del sistema di equazioni del modello.

Mentre le variabili endogene sono variabili stocastiche,³⁸ le variabili esogene non sono invece necessariamente di questo tipo, e possono essere sia probabilistiche che deterministiche. In ogni caso il loro valore è

³⁷ Per un approfondimento sulla stima dei valori dei “parametri strutturali” e degli “errori strutturali” si rimanda al Capitolo Quarto.

³⁸ Il termine *stocastico* viene utilizzato, in questo lavoro, in luogo del termine *casuale*, di uso più comune, per evitare confusioni terminologiche con l’attributo *causale*.

determinato al di fuori del modello, quindi non dipende né da alcuna variabile interna al modello, né dagli errori e .

Anche se l'interesse principale consiste nello stimare i valori dei parametri strutturali b che, esprimendo la forza dei nessi causali tra le variabili, legano le variabili dipendenti di ogni equazione alle rispettive indipendenti, la struttura di un modello di equazioni strutturali è definita da altri due insiemi di parametri: il primo, costituito dalle varianze e covarianze delle variabili esogene, il secondo dalle varianze e covarianze degli errori e .³⁹

3.4 La logica e le fasi di *Lisrel*.

Il *punto di partenza* di *Lisrel*, cioè il dato empirico da cui parte l'intero procedimento, è la matrice di varianza-covarianza fra le variabili osservate. Il *punto di arrivo* è costituito dai parametri di un modello di equazioni strutturali che descrivono i nessi causali fra le variabili.

La logica sottesa a quest'approccio ipotizza che, partendo da una certa relazione causale teorica, possiamo produrre una matrice di covarianza teorica che, confrontata con l'analoga matrice osservata, ci permetterà di capire quanto il nostro modello teorico è compatibile con i dati osservati.

In seguito vedremo che, lavorare non con i dati grezzi, ma con la matrice di varianza-covarianza fra le variabili osservate, semplifica di molto le operazioni connesse all'elaborazione.

Innanzitutto, si stabilisce a priori, su base puramente teorica e quindi pre-empirica, il modello causale. Ciò comporta la definizione di un certo numero di parametri, che diventano le incognite del modello da stimare. La loro stima avviene facendo interagire modello e dati, trovando cioè - a partire dai dati, ma con il vincolo del modello - quei valori dei parametri che, una volta collocati nel modello, producono lo scarto minore fra

³⁹ Per la trattazione di questo argomento si rimanda al Capitolo Quarto.

matrice di covarianza prodotta dal modello e matrice di covarianza osservata nei dati.

La verifica empirica di una teoria procede secondo tre fasi.

La prima fase è quella della *formulazione del modello*. Si tratta di tradurre la teoria in un sistema di equazioni strutturali, definendo le variabili osservate, ipotizzando le eventuali latenti, stabilendo i legami causali tra le variabili, e costruendo il modello complessivo in modo tale che esso possa essere matematicamente risolubile. Questa procedura porta, come punto conclusivo, alla definizione di un certo numero di parametri come entità incognite, ed alla eliminazione di altri possibili, ponendoli uguali a zero. Dunque, come risultato conclusivo di questa fase, abbiamo un certo numero di parametri da stimare.

La seconda fase è quella della *stima dei parametri strutturali* del modello mediante un processo iterativo di minimizzazione delle distanze fra i dati prodotti dal modello ed i dati osservati. Si parte attribuendo ai parametri dei valori iniziali più o meno arbitrari, si vede quale matrice di covarianza fra le variabili osservare produce questo modello, si misura la distanza di questa matrice “attesa” (cioè prodotta dal modello teorico) da quella reale “osservata” e, con procedure matematiche, si minimizza questa distanza, calcolando quali sarebbero i nuovi valori dei parametri per ridurre al minimo questa differenza. Il processo si chiude quando la distanza fra valori attesi e valori osservati non è ulteriormente riducibile mediante modifiche dei valori dei parametri. I parametri ottenuti sono, dunque, *i migliori possibili compatibili sia con i dati che con il modello*.

La terza fase di *Lisrel* è quella della *verifica del modello*, cioè del confronto fra modello teorico e dati osservati, per l’eventuale falsificazione del modello stesso.⁴⁰ Questa si basa sul confronto fra la matrice di

⁴⁰ La nozione di causalità come fondamento della scienza *empirica* è stata spesso messa in discussione in letteratura. La critica si basa sull’affermazione che la ricerca empirica può al massimo constatare la variazione simultanea fra due fenomeni, o il loro accadimento in successione temporale, ma ciò non è sufficiente per provare una relazione di causalità. Tuttavia, anche se una legge causale non può mai essere dimostrata empiricamente, essa è empiricamente saggibile. “*L’oggettività delle asserzioni della scienza risiede nel fatto che esse possono essere controllate intersoggettivamente*” [POPPER 1959; trad. it. 1970]. Il criterio che Popper propone è quello della *falsificabilità* (negativa), contrapposta alla *verificabilità* (positiva). Se non è possibile

covarianza “osservata” fra le variabili (quella ricavata dai dati) e la stessa matrice “attesa” (quella prodotta dal modello tramite i parametri stimati).

La distanza fra le due matrici è la minima compatibile con il modello, tuttavia può essere ancora troppo elevata per poter considerare il modello compatibile con i dati. Il residuo generato dalla discrepanza dati-modello è collegato all’esistenza dell’errore stocastico. Se la differenza è superiore a quella imputabile all’errore, il modello sarà respinto.

Se il modello è risultato inadeguato a descrivere i dati osservati, si apre quella che può essere definita la quarta fase di ***modifica del modello*** ed il ciclo della verifica ricomincia.

Le modifiche apportate si baseranno su valutazioni puramente teoriche, ma si avvarranno soprattutto dell’analisi del modello respinto.

3.5 La formulazione del modello secondo la notazione *Lisrel*.

La simbologia utilizzata da *Lisrel* per la formulazione del modello di equazioni strutturali può essere raggruppata in quattro categorie:

1. ***Variabili***. La distinzione base è quella fra variabili latenti e variabili osservate e fra variabili endogene e variabili esogene. Le variabili:

- ✓ *latenti endogene* si indicano con la lettera greca η (eta);
- ✓ *latenti esogene* si indicano con la lettera greca ξ (ksi);
- ✓ *osservate endogene* “ “ “ “ latina **Y**;
- ✓ *osservate esogene* “ “ “ “ latina **X**.

2. ***Errori stocastici***. L’errore:

accertare una teoria attraverso la ricerca empirica, tuttavia è possibile “falsificarla”, cioè respingerla sulla base del riscontro empirico. Dunque, se non è possibile provare che una teoria, o meglio, una relazione di causalità, è corretta, il confronto fra modello teorico e realtà osservata non avviene in positivo mediante la “prova” che il modello è esatto, ma si può realizzare solo in negativo, con la “non-falsificazione” del modello da parte dei dati, mediante la prova che i dati non contraddicono il modello. Per le sue caratteristiche, questo processo di verifica viene definito *valutazione dell’adattamento (fitting) del modello ai dati*. I *valori osservati* [valori generati dai dati] non coincideranno mai esattamente con i *valori attesi* [valori previsti dal modello], sia per il carattere semplificatorio del modello, sia per la presenza di un *errore stocastico* [errore casuale]. Lo scarto tra valori attesi e valori osservati viene definito *residuo* [dati = modello + residuo]; esso non può superare una certa soglia. E’ su questa soglia che si basa il processo di falsificazione del modello: se il residuo supera una certa soglia, allora il modello fornisce una spiegazione inadeguata della realtà e verrà respinto. [POPPER K. R., *The logic of scientific discovery*. New York, Basic Books, 1959; trad. it., *Logica della scoperta scientifica*. Torino, Einaudi, 1970].

- ✓ delle variabili η è indicato con la lettera greca ζ (zeta);
- ✓ “ “ Y “ “ “ “ ϵ (epsilon);
- ✓ “ “ X “ “ “ “ δ (delta).⁴¹

3. *Coefficienti strutturali (di regressione) fra le variabili legate da nessi causali.* I coefficienti:

- ✓ fra le variabili η e Y sono indicati con il simbolo λ^y (lambda-y);
- ✓ fra “ “ ξ e X “ “ “ “ λ^x (lambda-x);
- ✓ fra “ “ η e η “ “ la lettera greca β (beta);
- ✓ fra “ “ ξ e η “ “ “ “ γ (gamma).

4. *Varianze-covarianze fra le variabili esogene ξ (fra loro) e fra gli errori (fra errori dello stesso tipo).* Le varianze-covarianze:

- ✓ fra le variabili ξ sono indicate con la lettera greca ϕ (phi);
- ✓ fra gli errori ζ sono indicate con la lettera greca ψ (psi);
- ✓ fra gli errori ϵ sono indicate con il simbolo θ^ϵ (theta-epsilon);
- ✓ fra gli errori δ sono indicate con il simbolo θ^δ (theta-delta).

Tutti questi simboli hanno degli indici. Se si riferiscono ad una sola variabile (variabili ed errori), essi hanno un solo indice; se si riferiscono invece a due variabili (coefficienti di regressione e covarianze), essi hanno due indici.

3.6 La rappresentazione grafica del modello.

Nella rappresentazione grafica del modello di equazioni strutturali vengono riportate le variabili, i legami esistenti fra di esse ed i gli errori. Gli altri parametri del modello, cioè le covarianze delle variabili esogene e degli errori, non sempre vengono inseriti nel grafico.

I criteri stabiliti per la rappresentazione grafica di un modello secondo *Lisrel* sono i seguenti:

⁴¹ Si noti che gli errori ϵ e δ sono errori *nelle variabili X e Y* (o errori di misura), mentre gli errori ζ sono errori *nella equazione* e sono un aggregato di tutte le influenze su Y non esplicitate dal modello.

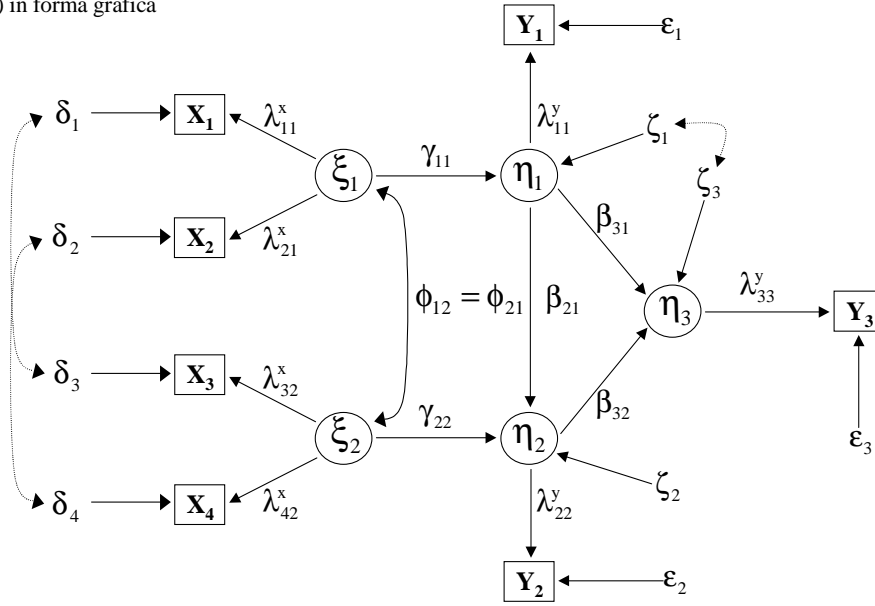
1. le variabili latenti sono racchiuse in un cerchio, mentre quelle osservate in un quadrato;
2. il legame causale diretto fra due variabili viene indicato con una freccia orientata che si dirige dalla variabile “causa” (indipendente), a quella “effetto” (dipendente). L’associazione (covariazione, correlazione) fra due variabili, senza che sia fornita un’interpretazione causale, viene indicata con una freccia a due direzioni che collega - con un tratto ad arco - le due variabili; l’assenza di frecce indica l’assenza di relazione fra due variabili;
3. la forza della relazione fra le due variabili implicate viene indicata riportando il valore del coefficiente relativo (di regressione se la freccia è orientata, cioè causale; coefficiente di correlazione o la covarianza se la freccia non ha direzione causale, cioè bidirezionale) in corrispondenza della freccia; l’assenza di tale valore sta ad indicare che il coefficiente è assunto pari a 1 (è il caso dei coefficienti fra errori e relative variabili dipendenti). Se il parametro strutturale è espresso in termini simbolici, esso presenterà due indici, il primo riferito alla variabile di arrivo della freccia (var. dipendente), il secondo alla variabile di partenza (var. indipendente); per le frecce curve bidirezionali, l’ordine degli indici sarà irrilevante.

In figura 3.6a) viene riportato un esempio di rappresentazione grafica di un modello di equazioni strutturali.

In questo modello sono presenti due variabili latenti esogene (ξ_1 e ξ_2) e tre variabili latenti endogene (η_1 , η_2 , η_3). Le relazioni fra queste variabili costituiscono il “nucleo” del modello, esprimendo le relazioni causali in esso presenti.

Figura 3.6: “Rappresentazione di un modello di equazioni strutturali secondo la notazione di *Lisrel*”.

a) in forma grafica



b) in forma algebrica

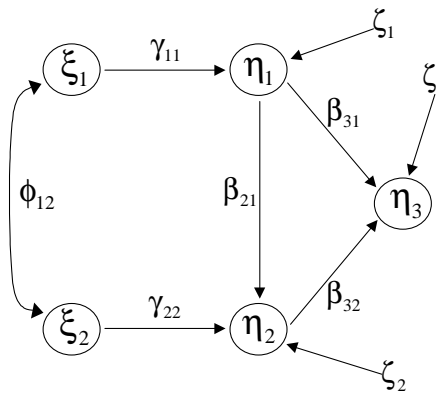
$$\begin{aligned}
 Y_1 &= \lambda_{11}^y \eta_1 + \varepsilon_1 \\
 Y_2 &= \lambda_{22}^y \eta_2 + \varepsilon_2 \\
 Y_3 &= \lambda_{33}^y \eta_3 + \varepsilon_3 \\
 \eta_1 &= \gamma_{11} \xi_1 + \zeta_1 \\
 \eta_2 &= \beta_{21} \eta_1 + \gamma_{22} \xi_2 + \zeta_2 \\
 \eta_3 &= \beta_{31} \eta_1 + \beta_{32} \eta_2 + \zeta_3 \\
 X_1 &= \lambda_{11}^x \xi_1 + \delta_1 \\
 X_2 &= \lambda_{21}^x \xi_1 + \delta_2 \\
 X_3 &= \lambda_{32}^x \xi_2 + \delta_3 \\
 X_4 &= \lambda_{42}^x \xi_2 + \delta_4
 \end{aligned}$$

Fonte: Corbetta P., *Metodi di analisi multivariata per le scienze sociali*. Bologna, Il Mulino, 1992.

Nell'esempio, le due variabili esogene ξ sono fra loro correlate, ma senza direzione causale (covarianza o coefficiente di correlazione: ϕ_{12}). Se ci fosse una direzione causale fra le due, non sarebbero più entrambe esogene, ma la dipendente sarebbe endogena.

Le variabili Y_1 , Y_2 e Y_3 sono gli indicatori, rispettivamente, delle variabili latenti (endogene) η_1 , η_2 , e η_3 . Le variabili X_1 e X_2 sono due indicatori della variabile latente (esogena) ξ_1 , mentre X_3 e X_4 sono due indicatori della variabile latente (esogena) ξ_2 . I coefficienti che legano le variabili latenti a quelle osservate sono i λ^y per il legame fra le η e le Y , ed i λ^x per il legame fra le ξ e le X .

Figura 3.7: “Parte del modello di fig. 3.6 relativa alla struttura di relazioni causali tra le variabili”.



Fonte: Corbetta P., *Metodi di analisi multivariata per le scienze sociali*. Bologna, Il Mulino, 1992.

Infine, nella rappresentazione grafica del modello, sono riportati gli errori stocastici, ognuno dei quali afferisce ad una variabile dipendente.

Dalla rappresentazione grafica esposta in figura 3.6a) è possibile ricavare il corrispondente sistema di equazioni strutturali, riportato nella parte b) della stessa figura. Ciascuna equazione presenta al primo membro la variabile dipendente, ed al secondo membro la somma di tanti addendi quante sono le variabili che agiscono causalmente sulla variabile dipendente, più l'errore stocastico.

Il modello *Lisrel* è costituito da tre parti, ciascuna riassumibile in una “equazione base”:

- il **modello strutturale**, per le relazioni causali fra le variabili endogene ed esogene;
- il **modello di misura** per la misurazione delle variabili endogene;
- il **modello di misura** per la misurazione delle variabili esogene.

3.7 Il modello strutturale.

Prendiamo in analisi, in questo contesto, la struttura di *relazioni causali* esistente tra le variabili *latenti*; si tratta della parte “causale” del modello, contrapposta a quella “di misura” e viene graficamente descritta in figura 3.7.

Questa parte del modello può essere formulata matematicamente tramite tre equazioni, aventi come primo membro le variabili η ; i termini

pari a 0, corrispondenti ad assenza di relazione fra le variabili, sono resi espliciti.

$$\begin{aligned}\eta_1 &= 0\eta_1 + 0\eta_2 + 0\eta_3 + \gamma_{11}\xi_1 + 0\xi_2 + \zeta_1 \\ \eta_2 &= \beta_{21}\eta_1 + 0\eta_2 + 0\eta_3 + 0\xi_1 + \gamma_{22}\xi_2 + \zeta_2 \\ \eta_3 &= \beta_{31}\eta_1 + \beta_{32}\eta_2 + 0\eta_3 + 0\xi_1 + 0\xi_2 + \zeta_3\end{aligned}$$

Queste equazioni possono essere espresse anche in forma matriciale:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}_{(3 \times 1)} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix}_{(3 \times 3)} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}_{(3 \times 1)} + \begin{bmatrix} \gamma_{11} & 0 \\ 0 & \gamma_{22} \\ 0 & 0 \end{bmatrix}_{(3 \times 2)} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}_{(2 \times 1)} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{bmatrix}_{(3 \times 1)}$$

Nel caso più generale, in cui le variabili endogene siano m e quelle esogene n , possiamo scrivere:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \dots \\ \eta_m \end{bmatrix}_{(m \times 1)} = \begin{bmatrix} 0 & \beta_{12} & \beta_{13} & \dots \\ \beta_{21} & 0 & \beta_{23} & \dots \\ \beta_{31} & \beta_{32} & 0 & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & 0 \end{bmatrix}_{(m \times m)} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \dots \\ \eta_m \end{bmatrix}_{(m \times 1)} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \dots \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \gamma_{mn} \end{bmatrix}_{(m \times n)} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \dots \\ \xi_n \end{bmatrix}_{(n \times 1)} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \dots \\ \zeta_m \end{bmatrix}_{(m \times 1)}$$

La formulazione matriciale, in notazione compatta, è la seguente:

$$\underset{(m \times 1)}{\boldsymbol{\eta}} = \underset{(m \times m)(m \times 1)}{\mathbf{B}\boldsymbol{\eta}} + \underset{(m \times n)(n \times 1)}{\boldsymbol{\Gamma}\boldsymbol{\xi}} + \underset{(m \times 1)}{\boldsymbol{\zeta}}$$

Questa è la **prima equazione base** delle tre che costituiscono il modello *Lisrel*. In essa compaiono:

a) i tre vettori delle variabili endogene $\boldsymbol{\eta}$, esogene $\boldsymbol{\xi}$ e degli errori $\boldsymbol{\zeta}$. I vettori $\boldsymbol{\eta}$ e $\boldsymbol{\zeta}$ contengono m elementi, tanti quante sono le variabili endogene; il vettore $\boldsymbol{\xi}$ contiene n elementi, pari al numero di variabili esogene;

b) la matrice dei coefficienti strutturali fra le variabili endogene (\mathbf{B}) e quella dei coefficienti fra le variabili esogene e le endogene ($\boldsymbol{\Gamma}$). La matrice \mathbf{B} è una matrice quadrata $(m \times m)$ di dimensioni pari al numero delle variabili endogene $\boldsymbol{\eta}$. La diagonale è sempre costituita da tutti 0, in quanto ad essi corrispondono i coefficienti di regressione di ciascuna variabile con

se stessa. La matrice Γ è invece di ordine $m \times n$. Questa parte del modello, per essere completamente specificata, necessita di altre due matrici:

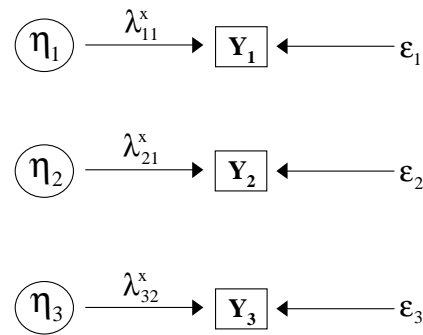
c) la matrice Φ che contiene le covarianze fra le variabili esogene ξ e la matrice Ψ che contiene le covarianze fra gli errori ζ . Trattandosi entrambe di matrici di covarianza, esse sono quadrate e simmetriche (quindi è possibile scriverle in forma triangolare). La matrice Φ è di ordine $n \times n$, dove n è il numero delle variabili esogene ξ . La matrice Ψ è di ordine $m \times m$, dove m è il numero delle variabili endogene η , e quindi degli errori ζ .

Le correlazioni esistenti fra le variabili esogene ξ , e cioè la loro matrice di covarianza Φ , è inclusa nel modello al fine di tener conto delle relazioni esistenti fra le variabili esplicitamente presenti. Il ruolo svolto dalla matrice Ψ di covarianza fra gli errori ζ è invece quello di consentire di includere nel modello l'effetto di variabili da questo escluse, ma invece operanti nella realtà dei dati osservati.

Nel caso in cui un modello venga perfettamente specificato, includendo tutte le variabili che effettivamente operano nella realtà, la componente "errore stocastico" di ciascuna equazione strutturale rappresenta una quantità trascurabile (che per assunzione ha media zero). In pratica, tuttavia, in questa componente verranno incluse anche tutte le variabili sconosciute che in realtà agiscono sulla variabile dipendente, ma che non sono presenti nel modello, in quanto non note o non misurabili. Se una di queste variabili sconosciute agisce contemporaneamente su due variabili endogene, ad esempio su η_1 e su η_3 (figura 3.7) e se non siamo a conoscenza di questo effetto, quindi non lo includiamo esplicitamente nel modello, il risultato sarà quello di ottenere una correlazione fra η_1 e η_3 che in realtà non esiste, cioè una correlazione spuria.⁴² Se invece includiamo esplicitamente nel modello una correlazione fra ζ_1 e ζ_3 [ponendo nella matrice Ψ il parametro $\psi_{31} \neq 0$] allora l'effetto congiunto della variabile

⁴² Cfr. paragrafo 3.2.3.

Figura 3.9: “Parte del modello di fig. 3.6 relativa alla misurazione delle variabili latenti endogene η ”.



Fonte: Corbetta P., *Metodi di analisi multivariata per le scienze sociali*. Bologna, Il Mulino, 1992.
 sconosciuta su η_1 e su η_3 sarà incluso nel modello e la correlazione spuria fra le due variabili non apparirà più.

In conclusione, dire che gli errori ζ_1 e ζ_3 sono correlati, significa ammettere l'esistenza di una variabile esterna sconosciuta che agisce contemporaneamente su η_1 e su η_3 . In figura 3.8 sono rappresentate le situazioni appena descritte.

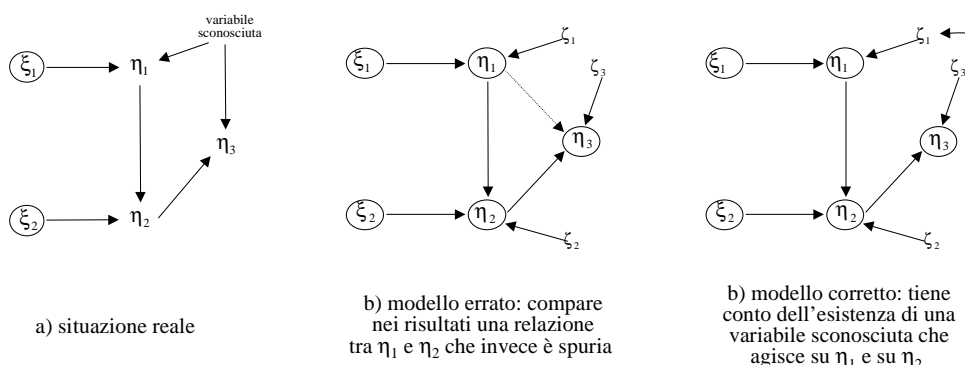
3.8 Il modello di misurazione per le variabili endogene.

Affrontiamo, in questa sede, il problema della misurazione dei legami esistenti tra le variabili *latenti* e le corrispondenti variabili *osservate*. Ci occupiamo dunque delle variabili endogene, i cui legami danno luogo alla seconda equazione base di *Lisrel*.

In figura 3.9 viene riportata la parte di modello di nostro interesse.

Le tre equazioni che definiscono questa parte del modello vengono riportate di seguito, in forma tale da esplicitare anche i termini aventi valore 0.

Figura 3.8: “Rappresentazione grafica dell'effetto operato dalla inclusione/esclusione nel modello della covarianza tra gli errori”.



$$Y_1 = \lambda_{11}^y + 0\eta_2 + 0\eta_3 + \varepsilon_1$$

$$Y_2 = 0\eta_1 + \lambda_{22}^y \eta_2 + 0\eta_3 + \varepsilon_2$$

$$Y_3 = 0\eta_1 + 0\eta_2 + \lambda_{33}^y \eta_3 + \varepsilon_3$$

Questo sistema di equazioni può essere espresso in forma matriciale nel modo seguente:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} \lambda_{11}^y & 0 & 0 \\ 0 & \lambda_{22}^y & 0 \\ 0 & 0 & \lambda_{33}^y \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

che, nel caso generale di p variabili osservate \mathbf{Y} e di m variabili latenti $\boldsymbol{\eta}$, assume la forma:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \dots \\ Y_p \end{bmatrix}_{(p \times 1)} = \begin{bmatrix} \lambda_{11}^y & \lambda_{12}^y & \lambda_{13}^y & \dots \\ \lambda_{21}^y & \lambda_{22}^y & \lambda_{23}^y & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \lambda_{pm}^y \end{bmatrix}_{(p \times m)} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \dots \\ \eta_m \end{bmatrix}_{(m \times 1)} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \dots \\ \varepsilon_n \end{bmatrix}_{(p \times 1)}$$

In notazione compatta, la **seconda equazione base** del modello *Lisrel* è la seguente: $\mathbf{Y} = \mathbf{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon}$
 $\begin{matrix} (p \times 1) & (p \times m) & (m \times 1) & (p \times 1) \end{matrix}$

In essa sono presenti:

- i tre vettori delle variabili endogene osservate \mathbf{Y} , endogene latenti $\boldsymbol{\eta}$ e degli errori $\boldsymbol{\varepsilon}$. I vettori \mathbf{Y} e $\boldsymbol{\varepsilon}$ contengono p elementi (tante quante sono le variabili osservate \mathbf{Y}); il vettore $\boldsymbol{\eta}$ contiene m elementi (tante quante sono le variabili latenti $\boldsymbol{\eta}$);
- la matrice dei coefficienti strutturali fra le variabili osservate e le variabili latenti $\mathbf{\Lambda}_y$ di ordine $p \times m$;
- la matrice di covarianza fra gli errori $\boldsymbol{\varepsilon}$, $\boldsymbol{\Theta}_\varepsilon$, quadrata e simmetrica, di ordine $p \times p$ (p è il numero degli errori $\boldsymbol{\varepsilon}$, pari a quello delle variabili osservate \mathbf{Y}).

E' bene precisare che, in pratica, contrariamente a quanto esposto in figura 3.9, avremo più variabili osservate per ogni variabile latente, quindi,

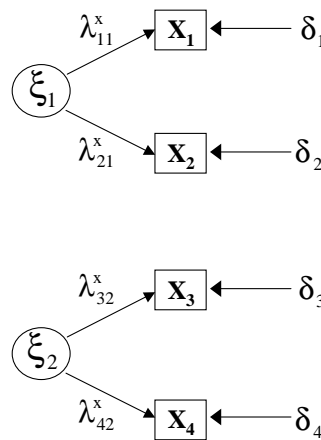
p (il numero delle \mathbf{Y}) sarà maggiore delle m (numero delle $\boldsymbol{\eta}$). Inoltre, non è detto che ci debba essere una corrispondenza biunivoca tra le $\boldsymbol{\eta}$ e le \mathbf{Y} . E' possibile che la stessa variabile latente agisca contemporaneamente su più variabili osservate. In altre parole, la matrice Λ_y è una matrice piena che può avere gli 0 distribuiti in maniera non sistematica.

3.9 Il modello di misurazione per le variabili esogene.

Le caratteristiche di questa parte del modello sono del tutto analoghe a quelle della parte precedente. Si tratta ancora di un modello di misurazione, avente questa volta per oggetto le variabili esogene.

La figura 3.10 mostra la parte del modello di nostro interesse.

Figura 3.10: “Parte del modello di fig. 3.6 relativa alla misurazione delle variabili latenti esogene ξ ”.



Fonte: Corbetta P., *Metodi di analisi multivariata per le scienze sociali*. Bologna, Il Mulino, 1992.

Le equazioni che possono essere desunte da questa parte del modello, in forma estesa che comprende anche i termini pari a zero, sono le seguenti:

$$X_1 = \lambda_{11}^x \xi_1 + 0\xi_2 + \delta_1$$

$$X_2 = \lambda_{21}^x \xi_1 + 0\xi_2 + \delta_2$$

$$X_3 = 0\xi_1 + \lambda_{32}^x \xi_2 + \delta_3$$

$$X_4 = 0\xi_1 + \lambda_{42}^x \xi_2 + \delta_4$$

In forma matriciale:

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} \lambda_{11}^x & 0 \\ \lambda_{21}^x & 0 \\ 0 & \lambda_{32}^x \\ 0 & \lambda_{42}^x \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix}$$

che nel caso più generale di q variabili osservate \mathbf{X} e di n variabili latenti ξ diventa:

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \dots \\ X_q \end{bmatrix}_{(q \times 1)} = \begin{bmatrix} \lambda_{11}^x & \lambda_{12}^x & \lambda_{13}^x & \dots \\ \lambda_{21}^x & \lambda_{22}^x & \lambda_{23}^x & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \lambda_{qn}^x \end{bmatrix}_{(q \times n)} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \dots \\ \xi_n \end{bmatrix}_{(n \times 1)} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \dots \\ \delta_q \end{bmatrix}_{(q \times 1)}$$

la cui scrittura in termini compatti dà luogo alla **terza equazione base** di *Lisrel*:

$$\mathbf{X}_{(q \times 1)} = \mathbf{\Lambda}_x_{(q \times n)} \boldsymbol{\xi}_{(n \times 1)} + \boldsymbol{\delta}_{(q \times 1)}$$

Analogamente al caso precedente, in quest'equazione troviamo:

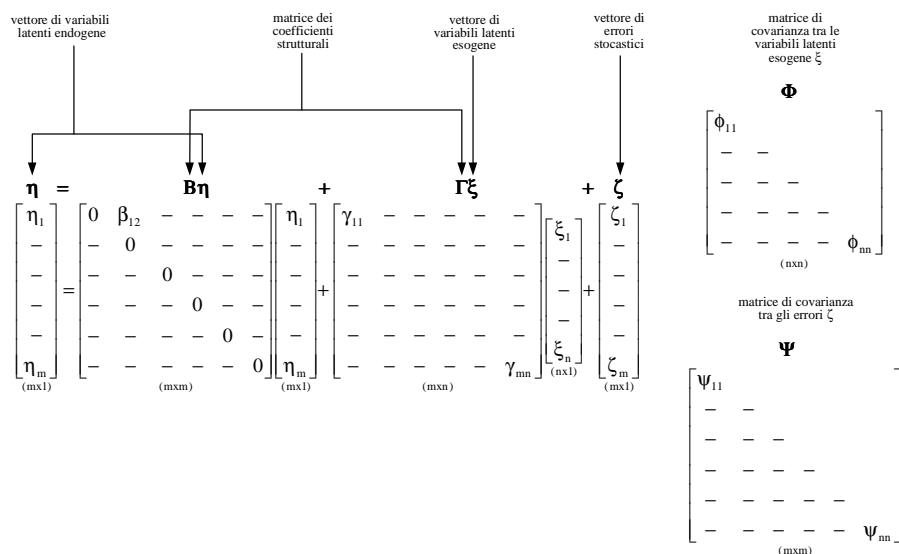
- i tre vettori delle variabili esogene osservate \mathbf{X} , esogene latenti ξ e degli errori δ . I vettori \mathbf{X} e $\boldsymbol{\delta}$ contengono q elementi (tante quante sono le variabili osservate \mathbf{X}); il vettore $\boldsymbol{\xi}$ contiene n elementi (tante quante sono le variabili latenti ξ);
- la matrice dei coefficienti strutturali fra le variabili osservate e quelle latenti $\mathbf{\Lambda}_x$ di ordine $q \times n$;
- la matrice di covarianza fra gli errori $\boldsymbol{\delta}$, $\boldsymbol{\Theta}_\delta$, quadrata e simmetrica, di ordine $q \times q$ (q è il numero degli errori $\boldsymbol{\delta}$, pari a quello delle variabili osservate \mathbf{X}).

La figura 3.11 mostra, in maniera riassuntiva, la rappresentazione completa del modello di equazioni strutturali in termini matriciali.

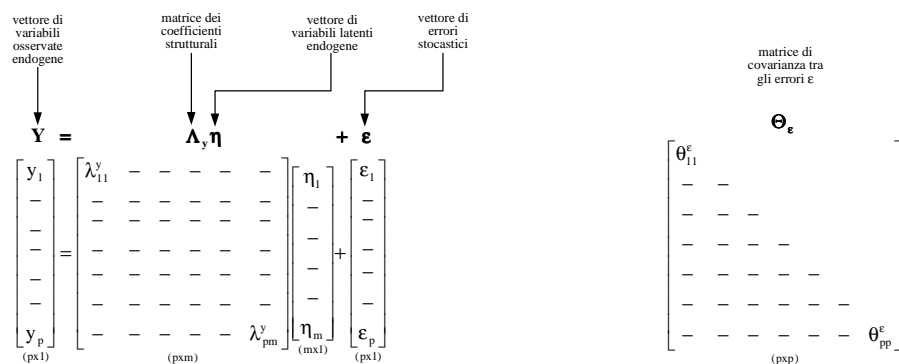
Per la specificazione completa del modello *Lisrel* sono necessarie, dunque, otto matrici: quattro di coefficienti strutturali (\mathbf{B} , $\mathbf{\Gamma}$, $\mathbf{\Lambda}_y$, $\mathbf{\Lambda}_x$), di forma generalmente rettangolare, e quattro matrici di covarianza (fra le variabili esogene: $\mathbf{\Phi}$; fra gli errori: $\mathbf{\Psi}$, $\mathbf{\Theta}_\epsilon$, $\mathbf{\Theta}_\delta$) anch'esse quadrate e simmetriche.

Figura 3.11: “Struttura generale del modello *Lisrel* (adattata da Hayduc [1987])”.

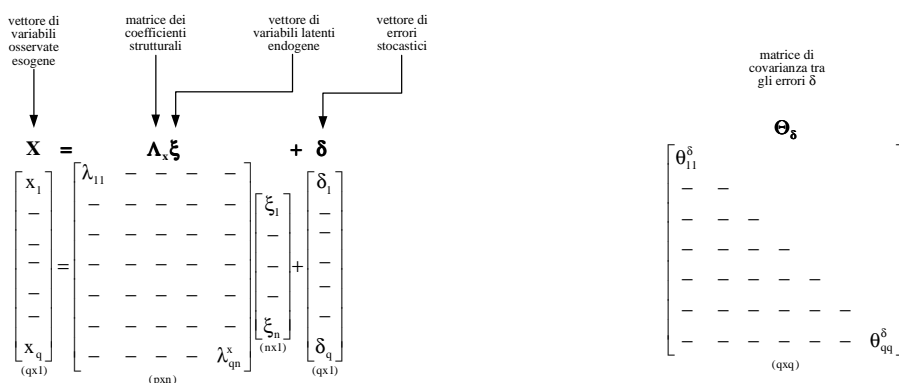
1. Il modello di equazioni strutturali



2. Il modello di misura per le variabili latenti endogene



3. Il modello di misura per le variabili latenti esogene



Fonte: Corbetta P., *Metodi di analisi multivariata per le scienze sociali*. Bologna, Il Mulino, 1992.

3.10 Le restrizioni del modello.

Le equazioni che abbiamo presentato poggiano sulle seguenti tre assunzioni:

- a) le variabili sono misurate in termini di scarti dalle loro medie, quindi:

$$E(\boldsymbol{\eta}) = E(\boldsymbol{\zeta}) = \mathbf{0}_{(m \times 1)}$$

$$E(\boldsymbol{\xi}) = \mathbf{0}_{(n \times 1)}$$

$$E(\mathbf{Y}) = E(\boldsymbol{\epsilon}) = \mathbf{0}_{(p \times 1)}$$

$$E(\mathbf{X}) = E(\boldsymbol{\delta}) = \mathbf{0}_{(q \times 1)}$$

- b) le variabili indipendenti e gli errori sono fra loro incorrelati; nella stessa equazione:

$$E(\boldsymbol{\xi}\boldsymbol{\zeta}') = \mathbf{0}_{(n \times m)}$$

$$E(\boldsymbol{\eta}\boldsymbol{\epsilon}') = \mathbf{0}_{(m \times p)}^{43}$$

$$E(\boldsymbol{\xi}\boldsymbol{\delta}') = \mathbf{0}_{(n \times q)}$$

fra equazioni:

$$E(\boldsymbol{\eta}\boldsymbol{\delta}') = \mathbf{0}_{(m \times q)}$$

$$E(\boldsymbol{\xi}\boldsymbol{\epsilon}') = \mathbf{0}_{(n \times p)}$$

⁴³ L'eguaglianza vale anche se le variabili sono invertite, in altre parole se $E(\boldsymbol{\xi}\boldsymbol{\zeta}') = \mathbf{0}_{(n \times m)}$, allora anche $E(\boldsymbol{\zeta}\boldsymbol{\xi}') = \mathbf{0}_{(m \times n)}$; l'osservazione vale anche per tutte le successive covarianze.

c) gli errori delle diverse equazioni sono fra loro incorrelati:

$$E(\zeta \epsilon') = \mathbf{0}_{(m \times p)}$$

$$E(\zeta \delta') = \mathbf{0}_{(m \times q)}$$

$$E(\epsilon \delta') = \mathbf{0}_{(p \times q)}$$

I punti b) e c) chiariscono che non sono possibili altri tipi di relazioni fra le variabili e gli errori oltre a quelle previste dalle otto matrici.

Ultima condizione da rispettare è la seguente:

d) nessuna delle equazioni strutturali deve essere ridondante, quindi, le equazioni del modello che esprimono le varie η devono essere fra loro indipendenti; ciò significa che nessuna variabile endogena η può essere una combinazione lineare di altre variabili endogene. In termini matriciali, la matrice \mathbf{B} è non singolare, cioè esiste \mathbf{B}^{-1} (\mathbf{B} è “positiva definita”).

CAPITOLO QUARTO

STIMA DEI PARAMETRI E VERIFICA EMPIRICA DEL MODELLO

AD EQUAZIONI STRUTTURALI

Sommario: - **4.1** La logica del procedimento di stima dei parametri del modello – **4.2** La covarianza fra le variabili esogene X - **4.3** La covarianza fra le variabili endogene Y - **4.4** La covarianza fra le variabili endogene Y e le esogene X – **4.5** La stima dei parametri strutturali - **4.6** L'adattamento del modello ai dati – **4.7** Le covarianze espresse in funzione dei parametri – **4.8** Regole di scomposizione delle covarianze e delle varianze – **4.9** Le misure di adattamento complessivo del modello – **4.10** Il miglioramento del modello – **4.10.1** Esclusione di parametri (*valori - t*) – **4.10.2** Inclusione di nuovi parametri (*indici di modifica*) – **4.11** Stima PLS

4.1 La logica del procedimento di stima dei parametri del modello.

Nei paragrafi precedenti abbiamo visto che il passaggio dalla rappresentazione diagrammatica del modello di equazioni strutturali a quella matriciale avviene assegnando ai parametri strutturali dei valori fissi (0 quando manca una relazione nel diagramma, 1, oppure altri valori), mentre gli altri parametri restano liberi e vanno stimati per poter valutare quantitativamente i nessi causali tra le variabili.

Punto di partenza per la stima dei parametri strutturali incogniti è la matrice di covarianza fra le variabili osservate **Y** e **X**. Innanzitutto, bisogna stabilire il legame algebrico fra il modello teorico e questa matrice; dato questo legame e noti i dati, arriveremo ad una stima dei valori numerici dei parametri strutturali.

La matrice di covarianza fra le variabili osservate \mathbf{Y} e \mathbf{X} può essere espressa in funzione delle otto matrici di parametri che caratterizzano il modello teorico [le quattro dei coefficienti strutturali, \mathbf{B} , $\mathbf{\Gamma}$, $\mathbf{\Lambda}_y$, $\mathbf{\Lambda}_x$, e le quattro matrici di covarianza, fra le variabili latenti esogene $\mathbf{\Phi}$ e fra gli errori $\mathbf{\Psi}$, $\mathbf{\Theta}_\epsilon$, $\mathbf{\Theta}_\delta$]. Dunque, se il modello teorico cambia, cambiando le configurazioni delle otto matrici, cambierà anche la matrice di covarianza “attesa” (cioè prodotta dal modello) fra le variabili osservate. Noto il nesso teorico che lega le otto matrici dei parametri strutturali alla matrice di covarianza “attesa” (o “teorica”) tra le variabili osservate, e nota la matrice di covarianza effettivamente “trovata” nei dati, è possibile procedere alla stima dei valori dei parametri che minimizzano la distanza fra queste due matrici.

Per dimostrare che la matrice di covarianza fra le variabili osservate può essere scritta in funzione delle otto matrici di parametri del modello, procederemo analizzando:

- la covarianza fra le variabili esogene (\mathbf{X});
- la covarianza fra le variabili endogene (\mathbf{Y});
- la covarianza fra le variabili esogene e le endogene.

Le dimostrazioni si basano su alcuni passaggi di algebra matriciale⁴⁴ e su un assunto fondamentale di *Lisrel*, cioè che in ognuna delle tre

⁴⁴ I passaggi algebrici delle dimostrazioni che seguono si basano sulle proprietà delle operazioni fra matrici e su quelle dell'operatore “valore atteso” E . In particolare, ricordiamo che:

$$\begin{aligned} E(k\mathbf{X}) &= kE(\mathbf{X}) \\ E(\mathbf{X} + \mathbf{Y}) &= E(\mathbf{X}) + E(\mathbf{Y}) \\ (\mathbf{A} + \mathbf{B})' &= \mathbf{A}' + \mathbf{B}' \\ (\mathbf{AB})' &= \mathbf{B}'\mathbf{A}' \end{aligned}$$

equazioni base del modello le variabili indipendenti sono incorrelate con gli errori.

4.2 La covarianza fra le variabili esogene \mathbf{X} .

La matrice di covarianza fra le variabili esogene \mathbf{X} può essere scritta come:

$$\left[\begin{array}{c} \text{Matrice di} \\ \text{covarianza} \\ \text{tra le } \mathbf{X} \end{array} \right] = \sum_{\mathbf{xx}} = \mathbf{E}(\mathbf{XX}')$$

Tenendo conto che $\mathbf{X} = \mathbf{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}$ (terza equazione base di *Lisrel*), otteniamo:

$$\begin{aligned} &= \mathbf{E} \left[(\mathbf{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta})(\mathbf{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta})' \right] \\ &= \mathbf{E} \left[(\mathbf{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta})(\boldsymbol{\xi}' \mathbf{\Lambda}_x' + \boldsymbol{\delta}') \right] \\ &= \mathbf{E} \left[\mathbf{\Lambda}_x \boldsymbol{\xi} \boldsymbol{\xi}' \mathbf{\Lambda}_x' + \boldsymbol{\delta} \boldsymbol{\xi}' \mathbf{\Lambda}_x' + \mathbf{\Lambda}_x \boldsymbol{\xi} \boldsymbol{\delta}' + \boldsymbol{\delta} \boldsymbol{\delta}' \right] \\ &= \mathbf{\Lambda}_x [\mathbf{E}(\boldsymbol{\xi} \boldsymbol{\xi}')] \mathbf{\Lambda}_x' + [\mathbf{E}(\boldsymbol{\delta} \boldsymbol{\xi}')] \mathbf{\Lambda}_x' + \mathbf{\Lambda}_x [\mathbf{E}(\boldsymbol{\xi} \boldsymbol{\delta}')] + \mathbf{E}(\boldsymbol{\delta} \boldsymbol{\delta}') \end{aligned}$$

Nella terminologia *Lisrel*: $\mathbf{E}(\boldsymbol{\xi} \boldsymbol{\xi}')$ è la matrice di covarianza tra le $\boldsymbol{\xi}$, cioè $\boldsymbol{\Phi}$; $\mathbf{E}(\boldsymbol{\delta} \boldsymbol{\delta}')$ è la matrice di covarianza tra gli errori $\boldsymbol{\delta}$, cioè $\boldsymbol{\Theta}_\delta$. Inoltre, per un assunto della terza equazione base di *Lisrel*, gli errori $\boldsymbol{\delta}$ sono incorrelati con le variabili indipendenti $\boldsymbol{\xi}$, vale a dire: $\mathbf{E}(\boldsymbol{\delta} \boldsymbol{\xi}') = \mathbf{0}_{(q \times n)}$ e

$$\mathbf{E}(\boldsymbol{\xi} \boldsymbol{\delta}') = \mathbf{0}_{(n \times q)}^{45}$$

[RIZZI A., *Il linguaggio delle matrici. Le applicazioni in economia, in statistica e nelle scienze sociali*. Roma, La Nuova Italia Scientifica, 1988; D'AMBRA L., *Lezioni di inferenza statistica*. Napoli, Rocco Curto Editore, 2000].

⁴⁵ Cfr. paragrafo 3.10, ipotesi b).

Possiamo scrivere, dunque:

$$\Sigma_{xx} = \Lambda_x \Phi \Lambda'_x + \Theta_\delta$$

che rappresenta l'equazione base dell'analisi fattoriale.

In questo modo, la matrice di covarianza delle variabili esogene \mathbf{X} viene espressa in funzione di una parte delle otto matrici di parametri *Lisrel*.

4.3 La covarianza fra le variabili endogene \mathbf{Y} .

Proseguendo in maniera analoga alla precedente, la matrice di covarianza delle variabili endogene \mathbf{Y} è la seguente:

$$\left[\begin{array}{c} \text{Matrice di} \\ \text{covarianza} \\ \text{tra le Y} \end{array} \right] = \Sigma_{yy} = E(\mathbf{Y}\mathbf{Y}')$$

Ricordando che $\mathbf{Y} = \Lambda_y \boldsymbol{\eta} + \boldsymbol{\varepsilon}$ (seconda equazione base di *Lisrel*), abbiamo:

$$\begin{aligned} &= E \left[(\Lambda_y \boldsymbol{\eta} + \boldsymbol{\varepsilon})(\Lambda_y \boldsymbol{\eta} + \boldsymbol{\varepsilon})' \right] \\ &= E \left[(\Lambda_y \boldsymbol{\eta} + \boldsymbol{\varepsilon})(\boldsymbol{\eta}' \Lambda_y' + \boldsymbol{\varepsilon}') \right] \\ &= E \left[\Lambda_y \boldsymbol{\eta} \boldsymbol{\eta}' \Lambda_y' + \boldsymbol{\varepsilon} \boldsymbol{\eta}' \Lambda_y' + \Lambda_y \boldsymbol{\eta} \boldsymbol{\varepsilon}' + \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}' \right] \\ &= \Lambda_y [E(\boldsymbol{\eta} \boldsymbol{\eta}')] \Lambda_y' + [E(\boldsymbol{\varepsilon} \boldsymbol{\eta}')] \Lambda_y' + \Lambda_y [E(\boldsymbol{\eta} \boldsymbol{\varepsilon}')] + E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}') \end{aligned}$$

Nella notazione *Lisrel*, $E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}')$ è la matrice di covarianza tra gli errori $\boldsymbol{\varepsilon}$, cioè Θ_ε ; sappiamo inoltre che, anche nella seconda equazione base, le variabili indipendenti $\boldsymbol{\eta}$ e gli errori $\boldsymbol{\varepsilon}$ sono incorrelati: $E(\boldsymbol{\varepsilon} \boldsymbol{\eta}') = \mathbf{0}_{(p \times m)}$ e

$E(\boldsymbol{\eta} \boldsymbol{\varepsilon}') = \mathbf{0}_{(m \times p)}$; ⁴⁶ per cui possiamo scrivere:

$$\Sigma_{yy} = \Lambda_y E(\boldsymbol{\eta} \boldsymbol{\eta}') \Lambda_y' + \Theta_\varepsilon$$

che è del tutto equivalente all'espressione della covarianza delle \mathbf{X} .

⁴⁶ Ibidem.

Per le variabili $\boldsymbol{\eta}$ abbiamo:

$$\begin{bmatrix} \text{Matrice di} \\ \text{covarianza} \\ \text{tra le } \boldsymbol{\eta} \end{bmatrix} = \sum_{\boldsymbol{\eta}\boldsymbol{\eta}} = E(\boldsymbol{\eta}\boldsymbol{\eta}')$$

Dalla prima equazione *Lisrel* sappiamo che $\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$, da cui, con alcune trasformazioni:

$$\begin{aligned} \boldsymbol{\eta} - \mathbf{B}\boldsymbol{\eta} &= \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \\ (\mathbf{I} - \mathbf{B})\boldsymbol{\eta} &= \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \\ \boldsymbol{\eta} &= (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}) \end{aligned}$$

Se l'inversa $(\mathbf{I} - \mathbf{B})^{-1}$ esiste (cioè nessuna variabile $\boldsymbol{\eta}$ è costituita da una combinazione lineare di altre variabili $\boldsymbol{\eta}$, quindi le equazioni del modello che esprimono le varie $\boldsymbol{\eta}$ sono indipendenti fra loro), è possibile inserire questa nuova formulazione di $\boldsymbol{\eta}$ nell'espressione della matrice di covarianza:

$$\begin{aligned} \sum_{\boldsymbol{\eta}\boldsymbol{\eta}} &= E(\boldsymbol{\eta}\boldsymbol{\eta}') \\ &= E\left(\left[(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta})\right]\left[(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta})\right]'\right) \\ &= E\left(\left[(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta})\right]\left[(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta})'(\mathbf{I} - \mathbf{B})^{-1'}\right]\right) \\ &= E\left[(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta})(\boldsymbol{\xi}'\boldsymbol{\Gamma}' + \boldsymbol{\zeta}')(\mathbf{I} - \mathbf{B})^{-1'}\right] \\ &= (\mathbf{I} - \mathbf{B})^{-1}E[\boldsymbol{\Gamma}\boldsymbol{\xi}\boldsymbol{\xi}'\boldsymbol{\Gamma}' + \boldsymbol{\zeta}\boldsymbol{\xi}'\boldsymbol{\Gamma}' + \boldsymbol{\Gamma}\boldsymbol{\xi}\boldsymbol{\zeta}' + \boldsymbol{\zeta}\boldsymbol{\zeta}'](\mathbf{I} - \mathbf{B})^{-1'} \\ &= (\mathbf{I} - \mathbf{B})^{-1}[\boldsymbol{\Gamma}(E(\boldsymbol{\xi}\boldsymbol{\xi}'))\boldsymbol{\Gamma}' + (E(\boldsymbol{\zeta}\boldsymbol{\xi}'))\boldsymbol{\Gamma}' + \boldsymbol{\Gamma}(E(\boldsymbol{\xi}\boldsymbol{\zeta}')) + E(\boldsymbol{\zeta}\boldsymbol{\zeta}')](\mathbf{I} - \mathbf{B})^{-1'} \\ &\text{Sapendo che } E(\boldsymbol{\xi}\boldsymbol{\xi}') = \text{matrice di covarianza fra le } \boldsymbol{\xi} = \boldsymbol{\Phi}; E(\boldsymbol{\zeta}\boldsymbol{\zeta}') = \\ &\text{matrice di covarianza tra gli errori } \boldsymbol{\zeta} = \boldsymbol{\Psi}; \text{ e ricordando che, nella prima} \end{aligned}$$

equazione base, gli errori ζ sono indipendenti dalle variabili latenti ξ , quindi, $E(\zeta\xi') = \mathbf{0}_{(m \times n)}$ e $E(\xi\zeta') = \mathbf{0}_{(n \times m)}$,⁴⁷ in conclusione:

$$\Sigma_{\eta\eta} = (\mathbf{I} - \mathbf{B})^{-1}(\mathbf{\Gamma}\mathbf{\Phi}\mathbf{\Gamma}' + \mathbf{\Psi})(\mathbf{I} - \mathbf{B})^{-1'}$$

Inserendo la nuova formulazione della matrice di covarianza tra le η nell'espressione della matrice di covarianza tra le \mathbf{Y} , otteniamo la notazione finale di tale matrice:

$$\Sigma_{yy} = \mathbf{\Lambda}_y \left[(\mathbf{I} - \mathbf{B})^{-1}(\mathbf{\Gamma}\mathbf{\Phi}\mathbf{\Gamma}' + \mathbf{\Psi})(\mathbf{I} - \mathbf{B})^{-1'} \right] \mathbf{\Lambda}_y' + \mathbf{\Theta}_\epsilon$$

4.4 La covarianza fra le variabili endogene \mathbf{Y} e le esogene \mathbf{X} .

Per esprimere la matrice di covarianza fra le \mathbf{X} e le \mathbf{Y} :

$$\left[\begin{array}{c} \text{Matrice di} \\ \text{covarianza} \\ \text{tra le X e le Y} \end{array} \right] = \Sigma_{xy} = E(\mathbf{X}\mathbf{Y}')$$

in funzione delle otto matrici di *Lisrel*, sostituiamo i valori di \mathbf{X} e di \mathbf{Y} che risultano dalla terza ($\mathbf{X} = \mathbf{\Lambda}_x\xi + \delta$) e dalla seconda equazione base ($\mathbf{Y} = \mathbf{\Lambda}_y\eta + \epsilon$); otteniamo così:

$$= E \left[(\mathbf{\Lambda}_x\xi + \delta)(\mathbf{\Lambda}_y\eta + \epsilon)' \right]$$

e, dopo alcuni passaggi algebrici:

$$\begin{aligned} &= E \left[(\mathbf{\Lambda}_x\xi + \delta)(\eta'\mathbf{\Lambda}_y' + \epsilon') \right] \\ &= E \left[\mathbf{\Lambda}_x\xi\eta'\mathbf{\Lambda}_y' + \delta\eta'\mathbf{\Lambda}_y' + \mathbf{\Lambda}_x\xi\epsilon' + \delta\epsilon' \right] \\ &= \mathbf{\Lambda}_xE(\xi\eta')\mathbf{\Lambda}_y' + E(\delta\eta')\mathbf{\Lambda}_y' + \mathbf{\Lambda}_xE(\xi\epsilon') + E(\delta\epsilon') \end{aligned}$$

Data l'assunzione che in ogni equazione base di *Lisrel* le variabili indipendenti e gli errori sono incorrelati, e che, inoltre, gli errori delle diverse equazioni sono incorrelati fra loro,⁴⁸ abbiamo:

⁴⁷ Ibidem.

⁴⁸ Cfr. paragrafo 3.10, ipotesi b) e c).

$$E(\delta\eta') = \mathbf{0}_{(q \times m)}; E(\xi\epsilon') = \mathbf{0}_{(n \times p)}; E(\delta\epsilon') = \mathbf{0}_{(q \times p)}$$

per cui:

$$\sum_{xy} = \Lambda_x E(\xi\eta') \Lambda_y'$$

Inserendo nel modello il valore di η si ottiene:

$$\begin{aligned} &= \Lambda_x E \left[\xi \left((\mathbf{I} - \mathbf{B})^{-1} (\Gamma\xi + \zeta) \right)' \right] \Lambda_y' \\ &= \Lambda_x E \left[\xi \left((\Gamma\xi + \zeta)' (\mathbf{I} - \mathbf{B})^{-1} \right)' \right] \Lambda_y' \\ &= \Lambda_x E \left[\xi \left((\xi'\Gamma' + \zeta') (\mathbf{I} - \mathbf{B})^{-1} \right)' \right] \Lambda_y' \\ &= \Lambda_x E \left[\xi \left(\xi'\Gamma' (\mathbf{I} - \mathbf{B})^{-1} + \zeta' (\mathbf{I} - \mathbf{B})^{-1} \right)' \right] \Lambda_y' \\ &= \Lambda_x E \left[\xi \xi'\Gamma' (\mathbf{I} - \mathbf{B})^{-1} + \xi \zeta' (\mathbf{I} - \mathbf{B})^{-1} \right] \Lambda_y' \\ &= \Lambda_x \left[E(\xi\xi') \Gamma' (\mathbf{I} - \mathbf{B})^{-1} + E(\xi\zeta') (\mathbf{I} - \mathbf{B})^{-1} \right] \Lambda_y' \end{aligned}$$

In conclusione, sapendo che $E(\xi\xi') =$ matrice di covarianza fra le $\xi = \Phi$, e che la prima equazione base assume assenza di correlazione fra le variabili indipendenti ξ e gli errori ζ , per cui $E(\xi\zeta') = \mathbf{0}_{(n \times m)}$,⁴⁹ l'espressione finale della matrice di covarianza fra le variabili osservate \mathbf{X} e \mathbf{Y} è la seguente:

$$\sum_{xy} = \Lambda_x \Phi \Gamma' (\mathbf{I} - \mathbf{B})^{-1} \Lambda_y'$$

L'espressione appena determinata indica che, noti i parametri strutturali del modello, è possibile calcolare la matrice di covarianza tra le variabili \mathbf{X} e \mathbf{Y} , quindi, *il modello implica una certa matrice di covarianza tra le variabili osservate.*

Anche se la dimostrazione complessiva si è sviluppata secondo tre linee separate, producendo tre equazioni conclusive, il ricercatore in genere

⁴⁹ Ibidem.

ha a disposizione, come punto di partenza e base empirica, la matrice di covarianza fra *tutte* le variabili osservate, senza distinzione fra le \mathbf{Y} e le \mathbf{X} .

E' quindi opportuno esprimere le tre equazioni:

$$\begin{aligned}\Sigma_{xx} &= \Lambda_x \Phi \Lambda'_x + \Theta_\delta \\ \Sigma_{yy} &= \Lambda_y \left[(\mathbf{I} - \mathbf{B})^{-1} (\Gamma \Phi \Gamma' + \Psi) (\mathbf{I} - \mathbf{B})^{-1} \right] \Lambda'_y + \Theta_\epsilon \\ \Sigma_{xy} &= \Lambda_x \Phi \Gamma' (\mathbf{I} - \mathbf{B})^{-1} \Lambda'_y\end{aligned}$$

nei termini di un'unica matrice di covarianza fra le variabili osservate. Si tratta della matrice *attesa stimata*, Σ , prodotta dal modello, a partire dalla stima dei parametri.

$$\Sigma = \begin{array}{c} \begin{array}{ccccccccc} Y_1 & Y_2 & Y_3 & \dots & Y_p & X_1 & X_2 & X_3 & \dots & X_q \end{array} \\ \left[\begin{array}{c|ccccccccc} Y_1 & & & & & & & & & \\ Y_2 & & & & & & & & & \\ Y_3 & & & & & & & & & \\ \dots & & & & & & & & & \\ \dots & & & & & & & & & \\ Y_p & & & & & & & & & \\ X_1 & & & & & & & & & \\ X_2 & & & & & & & & & \\ \dots & & & & & & & & & \\ X_q & & & & & & & & & \end{array} \right] \end{array}$$

Covarianze fra le Y
e le X (trasposta del
quadrante inferiore
sinistro)
Covarianze
fra le X e le Y
Covarianze fra le X

Sostituendo le espressioni algebriche determinate in precedenza, avremo:

$$\Sigma = \left[\begin{array}{c|ccccccccc} \Lambda_y \left[(\mathbf{I} - \mathbf{B})^{-1} (\Gamma \Phi \Gamma' + \Psi) (\mathbf{I} - \mathbf{B})^{-1} \right] \Lambda'_y + \Theta_\epsilon & & & & & & & & & \\ \hline \Lambda_x \Phi \Gamma' (\mathbf{I} - \mathbf{B})^{-1} \Lambda'_y & & & & & & & & & \end{array} \right]$$

(trasposta del quadrante
inferiore sinistro)
 $\Lambda_y (\mathbf{I} - \mathbf{B})^{-1} \Gamma \Phi \Lambda'_x$
 $\Lambda_x \Phi \Lambda'_x + \Theta_\delta$

La matrice di covarianza attesa Σ generata dal modello verrà confrontata con la matrice di covarianza \mathbf{S} osservata nei dati, al fine di valutare il buon adattamento del modello ai dati (*analisi dei residui*).

Se lo scarto $\mathbf{S} - \mathbf{\Sigma}$ è eccessivo, allora il modello non può essere considerato compatibile con i dati, perché la matrice di covarianza osservata nei dati e quella generata dal modello sono troppo distanti. Se invece, tale scarto può essere addebitato ad oscillazioni stocastiche, allora il modello non risulta falsificato, e pertanto non viene respinto.

4.5 La stima dei parametri strutturali.

La stima dei parametri strutturali, nel modello *Lisrel*, viene effettuata utilizzando *il metodo della massima verosimiglianza* (ML: *maximum likelihood*).

Siano x_1, \dots, x_n i valori di un campione casuale estratto da $X \sim f(x; \theta)$ e sia L la sua funzione di densità congiunta:

$$L(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

La funzione $L = L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$ è definita funzione di verosimiglianza e lo stimatore $\hat{\theta}$ è uno stimatore di massima verosimiglianza di θ , infatti, se sostituito nell'espressione precedente, la massimizza.

La funzione L dipende solo dal parametro θ incognito; per stimare tale parametro, si sceglie quel $\hat{\theta}$ tale che le osservazioni campionarie siano le più verosimili a rappresentare la distribuzione da cui sono estratte; in altri termini, si vuole determinare il parametro del modello statistico che verosimilmente abbia potuto generare i valori campionari ottenuti in via sperimentale.

Dalla definizione si evince che il metodo di massima verosimiglianza richiede la conoscenza della funzione di densità della popolazione; inoltre esso è applicabile se è possibile determinare il massimo della funzione di verosimiglianza, dunque, se la funzione è derivabile.

Con il metodo della massima verosimiglianza si esaminano come variano i valori della funzione di densità $f(x_1, x_2, \dots, x_n; \theta)$ per diversi valori di θ ; i valori campionari sono fissati (in quanto il campione è stato estratto), mentre il parametro θ varia nello spazio parametrico Θ .⁵⁰

Il criterio della massima verosimiglianza è un criterio generale che consente di stimare i parametri incogniti della popolazione, individuando quei parametri che generano la più elevata probabilità per i dati campionari di essere osservati.

Nel nostro caso specifico, esso consiste nell'individuare, data una certa matrice di covarianza osservata (in un campione) S , qual è la probabilità che questa matrice derivi da una certa matrice teorica Σ (nella popolazione); inoltre, permette di determinare quali valori attribuire ai parametri liberi nel modello che genera Σ , affinché la probabilità che S derivi da Σ sia la massima possibile.

Il processo di stima dei parametri strutturali “liberi” è iterativo, avviene cioè per stadi, per successive approssimazioni della stima ottimale.

Il punto di partenza è costituito dalle otto matrici di Lisrel, derivate dal modello teorico, che contengono dei parametri fissi, cioè aventi valori assegnati immodificabili, e dei parametri liberi, cioè incogniti da stimare.

In un primo stadio, si assegnano ai parametri liberi dei valori arbitrari. Sulla base della formula determinata nel paragrafo precedente, si calcola la matrice di covarianza Σ fra le variabili X e Y generata dal modello, con i parametri arbitrari. La matrice attesa Σ viene confrontata con l'analoga matrice S di covarianza tra le X e le Y osservata nei dati. Se le due matrici sono sufficientemente prossime, possiamo terminare l'analisi, concludendo che il modello non è falsificato dai dati.

In realtà, questo risultato non si ottiene mai al primo stadio del processo, specie se le stime iniziali sono arbitrarie.

⁵⁰ D'AMBRA L., *Lezioni di inferenza statistica*. Cit., pp. 141-142.

La matrice attesa Σ andrà migliorata apportando via via delle modifiche ai valori numerici dei parametri, fino a quando non si otterrà il massimo della prossimità tra Σ e S , cioè fino a quando qualsiasi modifica dei parametri porterà solo a peggioramenti della prossimità fra Σ e S . A questo punto, avremo ottenuto le migliori stime dei parametri compatibili con il modello. Dato il modello teorico, i valori numerici dei parametri non possono essere migliorati; se la distanza di Σ da S è ancora troppo alta, allora il modello teorico viene respinto, dunque, è «falsificato» dai dati.

Abbiamo osservato che il metodo della massima verosimiglianza ci consente di scegliere, fra tutti i possibili valori dei parametri liberi, quelli che generano una Σ la più prossima possibile a S , tale cioè che sia massima la probabilità che la S osservata nel campione derivi dalla Σ esistente nella popolazione.

Per poter proseguire, dobbiamo essere in grado di calcolare la probabilità di ottenere una certa S data una certa Σ . Questo è possibile attraverso la distribuzione di Wishart,⁵¹ che definisce tale probabilità e che ci consente, quindi, di definire un criterio per giudicare la prossimità tra Σ e S .

Per migliorare le stime dei parametri strutturali, massimizzando la prossimità fra Σ e S , sceglieremo quei valori che massimizzano la distribuzione di Wishart. Questa distribuzione è espressa in funzione dei parametri del modello: facendo le derivate parziali della funzione di Wishart rispetto ai parametri, troveremo quei valori numerici da attribuire loro per massimizzare la funzione, che rappresenta proprio la probabilità che S derivi dal modello Lisrel.

⁵¹ ANDERSON T.W., *An introduction to multivariate statistical analysis*. New York, Wiley, 1958.

4.6 L'adattamento del modello ai dati.

Caratteristica distintiva del modello ad equazioni strutturali, rispetto all'analisi fattoriale ed alla path analysis, è la possibilità di effettuare una fase di valutazione dell'adattamento del modello ai dati, prima di procedere alla determinazione dei parametri strutturali.

Una preliminare valutazione della soluzione trovata da Lisrel costituirà, in genere, il primo passo per giudicare il modello. E' infatti possibile che esso sia incompatibile con i dati immessi, o per errori nei dati, o perché il modello è radicalmente sbagliato.

Il test di ammissibilità viene effettuato alla decima iterazione (se il programma converge prima non viene effettuato); controlla che non ci siano delle varianze ≤ 0 e che le matrici Φ e Ψ siano positive definite. Se mancano questi requisiti, il programma si blocca e non procede con le iterazioni, né con le successive operazioni di calcolo dei residui, dei valori-t, degli indici di modifica, ecc..

Anche se il modello non risulta falsificato, e quindi non viene respinto, non è comunque possibile affermare che esso sia verificato. Sappiamo, infatti, che se un modello produce una ed una sola matrice di covarianza fra le variabili osservate, non è vero il contrario; gli stessi dati possono essere compatibili con molteplici modelli, ed il modello Lisrel potrebbe non essere necessariamente quello che effettivamente ha generato i dati osservati nella realtà.

In ogni caso, se il modello non viene falsificato, si procede all'analisi dei nessi causali tra le variabili; se, invece, esso risulta implausibile rispetto ai dati, allora questa fase non viene avviata.

Utilizzando tecniche tradizionali, quali l'analisi fattoriale e la path analysis, si giunge alla determinazione dei nessi causali tra le variabili, quindi, alla determinazione dei parametri, senza sottoporre il modello generale ad alcun test. L'attenzione è focalizzata sulla stima degli effetti, mentre nel

modello Lisrel è posta sulla plausibilità dell'intero modello. Solo se il test di ammissibilità è superato, si procede alla valutazione dei parametri stimati.

La fase di adattamento del modello ai dati (fitting) si basa sui residui del modello, ovvero sulla discrepanza fra la matrice di covarianza osservata, S , e la matrice attesa, Σ ; appare, quindi, necessario analizzare in maniera più approfondita il legame tra le covarianze fra le variabili osservate X e Y ed i parametri strutturali.

4.7 Le covarianze espresse in funzione dei parametri.

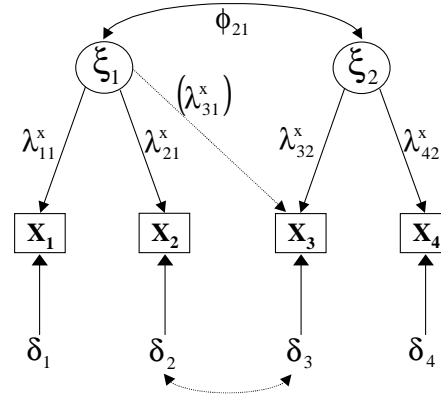
Per analizzare il legame esistente fra le covarianze delle variabili osservate ed i parametri del modello, consideriamo il modello in figura 4.1, dove, per semplicità espositiva, compaiono solo le variabili X e ξ ed i parametri λ_x , ϕ e $\theta\delta$ (modello di misurazione delle variabili latenti esogene).

In forma algebrica, questo modello è espresso dalle seguenti quattro equazioni:

$$\begin{array}{ll} X_1 = \lambda_{11}\xi_1 + \delta_1 & X_2 = \lambda_{21}\xi_1 + \delta_2 \\ X_3 = \lambda_{32}\xi_2 + \delta_3 & X_4 = \lambda_{42}\xi_2 + \delta_4 \end{array} \quad ^{52}$$

⁵² Per semplicità, poiché nel modello sono utilizzate solo le variabili esogene (osservate e latenti), indicheremo $\lambda^x = \lambda$ e $\theta\delta = \theta$.

Figura 4.1: “Modello con le sole variabili \mathbf{X} e ξ ”.



Fonte: Corbetta P., *Metodi di analisi multivariata per le scienze sociali*. Bologna, Il Mulino, 1992.

Esprimiamo le varianze e covarianze fra le variabili osservate \mathbf{X} , in termini di valore atteso. La varianza di X_1 è pari a:

$$\begin{aligned}
 \text{Var}(X_1) &= E(X_1 X_1) = \\
 &= E[(\lambda_{11} \xi_1 + \delta_1)(\lambda_{11} \xi_1 + \delta_1)] = \\
 &= E(\lambda_{11}^2 \xi_1 \xi_1 + 2\lambda_{11} \xi_1 \delta_1 + \delta_1 \delta_1) = \\
 &= \lambda_{11}^2 E(\xi_1 \xi_1) + 2\lambda_{11} E(\xi_1 \delta_1) + E(\delta_1 \delta_1) = \\
 &= \lambda_{11}^2 \text{Var}(\xi_1) + 2\lambda_{11} \text{Cov}(\xi_1 \delta_1) + \text{Var}(\delta_1)
 \end{aligned}$$

Poiché le covarianze fra le ξ e le δ sono nulle per definizione, utilizzando la notazione $\text{Var}(X_i) = \sigma_{ii}$, $\text{Var}(\xi_i) = \phi_{ii}$ e $\text{Var}(\delta_i) = \theta_{ii}$, abbiamo:

$$\sigma_{11} = \lambda_{11}^2 \phi_{11} + \theta_{11}$$

In modo analogo, calcoliamo la covarianza tra X_1 e X_2 :

$$\begin{aligned}
 \text{Cov}(X_1 X_2) &= E(X_1 X_2) = \\
 &= E[(\lambda_{11} \xi_1 + \delta_1)(\lambda_{21} \xi_1 + \delta_2)] = \\
 &= E(\lambda_{11} \lambda_{21} \xi_1 \xi_1 + \lambda_{11} \xi_1 \delta_2 + \lambda_{21} \xi_1 \delta_1 + \delta_1 \delta_2) = \\
 &= \lambda_{11} \lambda_{21} E(\xi_1 \xi_1) + \lambda_{11} E(\xi_1 \delta_2) + \lambda_{21} E(\xi_1 \delta_1) + E(\delta_1 \delta_2) = \\
 &= \lambda_{11} \lambda_{21} \text{Var}(\xi_1) + \lambda_{11} \text{Cov}(\xi_1 \delta_2) + \lambda_{21} \text{Cov}(\xi_1 \delta_1) + \text{Cov}(\delta_1 \delta_2)
 \end{aligned}$$

Ricordando che $\text{Cov}(\xi_1 \delta_2) = \text{Cov}(\xi_1 \delta_1) = 0$ per definizione e che $\text{Cov}(\delta_1 \delta_2) = 0$ per costruzione del modello:

$$\sigma_{21} = \lambda_{11} \lambda_{21} \phi_{11}$$

Infine, la covarianza tra X_1 e X_3 è pari a:

$$\begin{aligned} \text{Cov}(X_1 X_3) &= E(X_1 X_3) = \\ &= E[(\lambda_{11} \xi_1 + \delta_1)(\lambda_{32} \xi_2 + \delta_3)] = \\ &= E(\lambda_{11} \lambda_{32} \xi_1 \xi_2 + \lambda_{11} \xi_1 \delta_3 + \lambda_{32} \xi_2 \delta_1 + \delta_1 \delta_3) = \\ &= \lambda_{11} \lambda_{32} E(\xi_1 \xi_2) + \lambda_{11} E(\xi_1 \delta_3) + \lambda_{32} E(\xi_2 \delta_1) + E(\delta_1 \delta_3) = \\ &= \lambda_{11} \lambda_{32} \text{Cov}(\xi_1 \xi_2) + \lambda_{11} \text{Cov}(\xi_1 \delta_3) + \lambda_{32} \text{Cov}(\xi_2 \delta_1) + \text{Cov}(\delta_1 \delta_3) \end{aligned}$$

In forma compatta:

$$\sigma_{31} = \lambda_{11} \lambda_{32} \phi_{21}$$

Le quattro equazioni delle variabili \mathbf{X} vengono indicate in forma matriciale con la notazione:

$$\mathbf{X} = \Lambda_x \boldsymbol{\xi} + \boldsymbol{\delta}$$

cioè:

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ 0 & \lambda_{32} \\ 0 & \lambda_{42} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix}$$

Come abbiamo dimostrato nel paragrafo 4.2, la matrice di covarianza fra le \mathbf{X} è pari a:

$$\Sigma_{xx} = \Lambda_x \Phi \Lambda'_x + \Theta_\delta$$

che nel nostro caso assume la seguente

forma:

$$\begin{aligned} \Sigma_{xx} &= \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ 0 & \lambda_{32} \\ 0 & \lambda_{42} \end{bmatrix} \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \lambda_{11} & \lambda_{12} & 0 & 0 \\ 0 & 0 & \lambda_{32} & \lambda_{42} \end{bmatrix} + \begin{bmatrix} \theta_{11} & 0 & 0 & 0 \\ 0 & \theta_{22} & 0 & 0 \\ 0 & 0 & \theta_{33} & 0 \\ 0 & 0 & 0 & \theta_{44} \end{bmatrix} \\ &= \begin{bmatrix} \lambda_{11}^2 \phi_{11} + \theta_{11} & \lambda_{11} \lambda_{21} \phi_{11} & \lambda_{11} \lambda_{32} \phi_{21} & \lambda_{11} \lambda_{42} \phi_{21} \\ \lambda_{21} \lambda_{11} \phi_{11} & \lambda_{21}^2 \phi_{11} + \theta_{22} & \lambda_{21} \lambda_{32} \phi_{21} & \lambda_{21} \lambda_{42} \phi_{21} \\ \lambda_{32} \lambda_{11} \phi_{21} & \lambda_{32} \lambda_{21} \phi_{21} & \lambda_{32}^2 \phi_{22} + \theta_{33} & \lambda_{32} \lambda_{42} \phi_{22} \\ \lambda_{42} \lambda_{11} \phi_{21} & \lambda_{42} \lambda_{21} \phi_{21} & \lambda_{42} \lambda_{32} \phi_{22} & \lambda_{42}^2 \phi_{22} + \theta_{44} \end{bmatrix} \end{aligned}$$

Se le variabili ξ sono standardizzate, quindi $\phi_{11} = \phi_{22} = 1$, abbiamo:

$$\Sigma_{xx} = \begin{bmatrix} \lambda_{11}^2 + \theta_{11} & \lambda_{11} \lambda_{21} & \lambda_{11} \lambda_{32} \phi_{21} & \lambda_{11} \lambda_{42} \phi_{21} \\ \lambda_{21} \lambda_{11} & \lambda_{21}^2 + \theta_{22} & \lambda_{21} \lambda_{32} \phi_{21} & \lambda_{21} \lambda_{42} \phi_{21} \\ \lambda_{32} \lambda_{11} \phi_{21} & \lambda_{32} \lambda_{21} \phi_{21} & \lambda_{32}^2 + \theta_{33} & \lambda_{32} \lambda_{42} \\ \lambda_{42} \lambda_{11} \phi_{21} & \lambda_{42} \lambda_{21} \phi_{21} & \lambda_{42} \lambda_{32} & \lambda_{42}^2 + \theta_{44} \end{bmatrix}$$

4.8 Regole di scomposizione delle covarianze e delle varianze.

Dal modello semplificato presentato nel paragrafo precedente è possibile ricostruire le due **regole di scomposizione delle covarianze e delle varianze** fra le variabili, tenendo presente che tale formulazione è valida solo per variabili standardizzate.⁵³

Prima regola di scomposizione (scomposizione delle correlazioni): la correlazione fra due variabili può essere scomposta in tanti addendi quanti sono i percorsi che le collegano;⁵⁴ ogni addendo è dato dal prodotto dei coefficienti incontrati sul percorso.

⁵³ Lavorando con variabili standardizzate, d'ora in avanti, sarà più corretto parlare di *correlazioni* anziché di *covarianze*.

⁵⁴ Utilizzando la terminologia della *path analysis*, si definisce **percorso** "il tracciato fatto di sequenze e di frecce direzionali che unisce due variabili". [CORBETTA P., *Metodi di analisi multivariata per le scienze sociali*. Cit., p. 122].

Considerando ancora, a titolo di esempio, il modello in figura 4.1, le variabili X_1 e X_2 sono collegate dal percorso:

$$X_1 \leftarrow \xi_1 \rightarrow X_2$$

La loro correlazione sarà:

$$\sigma_{21} = \lambda_{11}\lambda_{21}$$

I coefficienti del percorso possono essere i coefficienti di regressione, se il legame ha una direzione, oppure i coefficienti di correlazione, se il legame è bidirezionale. Nel caso delle variabili X_2 e X_3 il percorso è:

$$X_2 \leftarrow \xi_1 \leftrightarrow \xi_2 \rightarrow X_3$$

La loro correlazione sarà:

$$\sigma_{32} = \lambda_{21}\phi_{21}\lambda_{32}$$

Se esistesse un legame fra ξ_1 e X_3 (in figura 4.1 indicato con la freccia tratteggiata), allora ci sarebbero due percorsi fra X_2 e X_3 :

$$X_2 \leftarrow \xi_1 \leftrightarrow \xi_2 \rightarrow X_3$$

$$X_2 \leftarrow \xi_1 \rightarrow X_3$$

e quindi la correlazione sarebbe la somma di due addendi:

$$\sigma_{32} = \lambda_{21}\phi_{21}\lambda_{32} + \lambda_{21}\lambda_{31}$$

Seconda regola di scomposizione (scomposizione delle varianze): la varianza di una variabile dipendente si scompone in varianza spiegata dalle variabili agenti causalmente su di essa ed in varianza non spiegata.

La varianza spiegata è data da tanti addendi quante sono le variabili agenti causalmente in modo diretto su quella variabile; ogni addendo rappresenta il legame diretto (coefficiente di regressione) fra le due variabili, moltiplicato per il coefficiente di correlazione fra le due variabili (che, per la prima regola di scomposizione, definisce tutti i legami diretti ed indiretti fra le due variabili). La varianza non spiegata è la varianza dell'errore.

Dall'esempio in figura 4.1, la varianza di X_3 è pari a:

$$\underset{\substack{\text{varianza} \\ \text{complessiva}}}{\sigma_{33}} = \underset{\substack{\text{varianza} \\ \text{spiegata}}}{\lambda_{32}^2} + \underset{\substack{\text{varianza} \\ \text{non spiegata}}}{\theta_{33}}$$

cioè,

λ_{32} (legame diretto fra X_3 e la variabile ξ_2 , unica agente casualmente in modo diretto su di essa);

×

λ_{32} (tutti i legami diretti ed indiretti fra le due variabili, in questo caso uno solo);

+

θ_{33} (varianza non spiegata = varianza dell'errore δ_3).

Se il legame fra ξ_1 e X_3 fosse presente, avremmo:

$$\sigma_{33} = \lambda_{31}(\lambda_{31} + \lambda_{32}\phi_{21}) + \lambda_{32}(\lambda_{32} + \lambda_{31}\phi_{21}) + \theta_{33}$$

Per la regola di scomposizione delle correlazioni potremmo anche scrivere:

$$\sigma_{33} = \lambda_{31}\sigma_{X_3\xi_1} + \lambda_{32}\sigma_{X_3\xi_2}$$

La varianza spiegata è pari al coefficiente di determinazione (cioè al quadrato del coefficiente di correlazione multipla R^2) dell'equazione di regressione relativa alla variabile considerata. Dunque, se le variabili sono standardizzate, la varianza non spiegata sarà pari a $1-R^2$.

$$\underset{\substack{\text{varianza} \\ \text{complessiva}}}{\sigma_{11}} = 1 = \underset{\substack{\text{varianza} \\ \text{spiegata}}}{R^2} + \underset{\substack{\text{varianza} \\ \text{non spiegata}}}{\text{Var}(e_1)}$$

La prima regola di scomposizione è estremamente importante, perché consente di capire l'utilizzazione dell'analisi dei residui al fine di migliorare la specificazione del modello. Essa, infatti, instaura un legame non solo algebrico, ma anche graficamente visualizzabile, fra i parametri del modello e le covarianze delle variabili osservate.

Se, ad esempio, fra X_2 e X_3 risulta un alto residuo, ciò significa che la covarianza osservata s_{32} è maggiore di quella stimata dal modello σ_{32} , vale a dire che i legami previsti dal modello fra le due variabili sono

insufficienti. Si potrebbe allora pensare di inserire un legame diretto fra ξ_1 e X_3 (tratteggiato in figura), che rafforzi i legami fra X_2 e X_3 , oppure di inserire un legame esterno che leghi X_2 a X_3 , espresso da un legame fra i loro errori.

Se le variabili con alti residui fossero più di due, potrebbe essere necessario inserire una variabile latente che agisca su di esse.

In generale, se il modello sottoposto a stima non include tutti i percorsi effettivamente esistenti fra due variabili i e j , allora la covarianza σ_{ij} stimata dal modello risulta inferiore a quella osservata s_{ij} , e le due variabili presentano un alto residuo positivo ($s_{ij} - \sigma_{ij} > 0$). Di conseguenza, gli altri percorsi fra i e j inclusi nel modello hanno dei parametri sovradimensionati rispetto a quelli reali, in quanto la procedura di stima cerca di avvicinare il più possibile le covarianze stimate a quelle osservate. Mancando nel modello alcuni dei percorsi fra i e j , parte della covarianza fra queste variabili verrà assorbita dai percorsi esistenti, con la sovrastima dei rispettivi parametri.

Durante la fase di miglioramento del modello, dopo aver individuato i residui elevati, è possibile procedere alla loro eliminazione:

- a) introducendo fra le variabili già presenti nel modello dei legami aggiuntivi che, in modo diretto o indiretto, coinvolgano le variabili con residuo elevato;
- b) introducendo delle nuove variabili latenti che agiscano su quelle affette da residui elevati;
- c) introducendo dei legami fra gli errori delle variabili dai residui elevati.

4.9 Le misure di adattamento complessivo del modello.

Le misure di adattamento del modello ai dati sono tutte funzioni del residuo, cioè dello scarto fra S e Σ . Questo deve essere formulato nei termini di una distribuzione statistica nota, in modo da poter prescindere,

nel confronto fra i due valori di \mathbf{S} e di $\mathbf{\Sigma}$, dalle oscillazioni stocastiche di campionamento. Se tale distribuzione è nota, data una certa differenza fra \mathbf{S} e $\mathbf{\Sigma}$, sarà possibile distinguere quanto di tale differenza è dovuto alle oscillazioni stocastiche e quanto invece ad una reale discrepanza fra le due matrici.

Se il modello è corretto ed il campione sufficientemente grande, è possibile dimostrare che la funzione di adattamento del modello (*Statistica T*) – che è una funzione dello scarto ($\mathbf{S} - \mathbf{\Sigma}$) ed è posta alla base della stima dei parametri del modello tramite la sua minimizzazione – si distribuisce secondo la distribuzione del χ^2 , con df gradi di libertà, dove df è pari a:

$$df = \frac{1}{2} (p + q)(p + q + 1) - t$$

in cui t è il numero di parametri *liberi* da stimare, p il numero delle variabili \mathbf{Y} e q il numero delle variabili \mathbf{X} .

Per costruzione, df è pari al numero di varianze-covarianze in input, meno il numero di parametri *liberi* da stimare; questa differenza rappresenta il numero di parametri *fissi*.

Lisrel calcola il valore T della funzione dello scarto fra \mathbf{S} e $\mathbf{\Sigma}$; esso viene confrontato, sulle tavole della distribuzione del χ^2 , con il valore confrontato corrispondente, per lo stesso grado di libertà, ad una probabilità $P = 0,10$.

L'ipotesi nulla da sottoporre a test è quella che il modello sia «vero», cioè quella dell'identità fra \mathbf{S} e $\mathbf{\Sigma}$ ($\mathbf{S} - \mathbf{\Sigma} = \mathbf{0}$); se il valore di T trovato è inferiore al valore tabulato, allora l'ipotesi nulla non è falsificata ed il modello teorico non è respinto. In questo caso è possibile affermare che lo scarto $\mathbf{S} - \mathbf{\Sigma}$ è sufficientemente piccolo da poter essere attribuito a fluttuazioni stocastiche, e non ad una reale differenza fra le due matrici. Se invece il valore ottenuto è inferiore a quello tabulato, allora l'ipotesi nulla dell'identità fra \mathbf{S} e $\mathbf{\Sigma}$ deve essere respinta.

Oltre a valutare *in assoluto* l'adeguatezza del modello, il test chi-quadrato è utilizzato *in relativo* per mettere a confronto due modelli, uno dei quali contiene solo una parte dei parametri dell'altro.

Dato, ad esempio, un modello con un certo valore della statistica T , sulla base di opportune considerazioni, si fissano determinati parametri, cioè si escludono dal modello eguagliandoli a zero; dopo questa operazione, si stima di nuovo il modello, che avrà T e df superiori al modello precedente.

Per valutare la crescita di T si controlla sulle tavole del χ^2 la significatività della differenza fra i T dei due modelli, con un numero di gradi di libertà corrispondente alla differenza fra i due gradi di libertà; se la differenza è non-significativa, i parametri esclusi per prova possono essere lasciati fuori dall'analisi e, quindi, il nuovo modello è preferibile al precedente, in quanto è più parsimonioso, senza manifestare un significativo peggioramento dell'adattamento ai dati.

Maggiori sono i gradi di libertà del modello, minori sono i parametri di cui esso ha bisogno per esprimere la struttura delle covarianze (parametri da stimare); dunque, il modello ha una maggiore capacità di semplificazione della realtà.

Ai fini della valutazione complessiva, fra due modelli con T dello stesso livello di significatività, ma con gradi di libertà differenti, si sceglierà quello più parsimonioso, più semplice, cioè quello con maggiori gradi di libertà. In questo caso, anziché utilizzare il valore T come misura di valutazione complessiva del modello, si costruisce il rapporto:

$$\frac{T}{df} = \text{indice di bontà del modello}$$

che tiene conto non solo dell'adattamento fra \mathbf{S} e Σ , ma anche della parsimoniosità del modello.

Secondo alcuni autori,⁵⁵ un rapporto T/df compreso fra 1 e 3 è accettabile per la non falsificazione del modello.

Tutte le statistiche che fanno riferimento alla distribuzione chi-quadrato hanno come limite la loro sensibilità alla dimensione del campione, perché il valore della statistica aumenta proporzionalmente all'aumentare di N . Nel nostro caso, per campioni con N elevato, è frequente che la statistica T assuma valori significativi (corrispondenti a falsificazione del modello) anche in situazioni di buon adattamento del modello ai dati. Inoltre, risulta difficoltoso confrontare fra loro statistiche T provenienti da campioni di diversa numerosità.

Per ovviare a questi limiti, è possibile costruire delle misure alternative di adattamento generale del modello. La prima è il ***goodness of fit index*** (GFI), dove il valore della statistica T viene standardizzato con il valore massimo che essa può raggiungere:

$$0 \leq \left\{ \text{GFI} = 1 - \frac{T_i}{\max(T_i)} \right\} \leq 1$$

Questa misura assume valori compresi fra 0 (pessimo adattamento del modello ai dati) ed 1 (perfetto adattamento). Essa risulta di facile interpretazione, inoltre dà la possibilità di confrontare modelli su insiemi diversi di dati, ma non tiene conto dei gradi di libertà e, quindi, della parsimoniosità del modello. Una versione modificata di tale indice è l'***adjusted goodness of fit index*** (AGFI), così definito:

$$0 \leq \left\{ \text{AGFI} = 1 - \left(\frac{k}{df} \right) (1 - \text{GFI}) \right\} \leq 1$$

dove df sono i gradi di libertà e k il numero di varianze-covarianze in input, pari a $\frac{1}{2}(p+q)(p+q+1)$.

⁵⁵ CARMINES E., MCIVER J., *Analyzing models with unobserved variables: analysis of covariance structures*. "Social Measurement", Beverly Hills, Sage.

I due indici presentano lo svantaggio di avere una distribuzione statistica non nota, per cui non è possibile utilizzarli per effettuare un test di significatività del modello.

L'ultima misura calcolata da *Lisrel* è la **Root mean squared residuals** (RMR: la radice quadrata della media dei residui al quadrato), così definita:

$$0 \leq \text{RMR} = \sqrt{\frac{1}{k} \sum (s_{ij} - \sigma_{ij})^2}$$

Quest'indice è nullo quando S coincide con Σ , ma, a differenza dei precedenti, non ha soglia superiore. Presenta, quindi, gli stessi limiti della statistica T , per cui serve solo per confrontare diversi modelli, ma calcolati sugli stessi dati; inoltre, anche la sua distribuzione statistica non è nota. La RMR, tuttavia, non dipende dalla numerosità campionaria, per cui, in presenza di campioni numerosi, può essere più adatta della statistica T per valutare il modello.

4.10 Il miglioramento del modello.

Dopo aver calcolato le misure di adattamento *complessivo* del modello ai dati osservati, si procede con la fase di miglioramento *analitico*, operando sui singoli parametri o singoli residui del modello. Il fine ultimo è comunque quello di pervenire ad un miglioramento complessivo del modello stesso.

Questo processo di miglioramento analitico può avvenire secondo due modalità:

- esclusione di parametri dal modello;
- inclusione di nuovi parametri;

4.10.1 Esclusione di parametri (*valori - t*).

I valori dei parametri prodotti da *Lisrel* sono delle stime campionarie dei valori effettivi e, per tale ragione, sono affetti da oscillazioni stocastiche. Il controllo di significatività dei singoli parametri è finalizzato ad eliminare quelli non significativamente diversi da 0, ossia quelli che assumono un valore diverso da 0 per effetto di oscillazioni stocastiche, ma per i quali non è possibile respingere l'ipotesi che assumano valore 0 nell'universo.

Nel caso di distribuzione campionaria normale, si sottopone a verifica l'ipotesi nulla $\mathbf{H}_0: \mathbf{P} = \mathbf{0}$, contro l'ipotesi alternativa $\mathbf{H}_1: \mathbf{P} \neq \mathbf{0}$ (il parametro P è significativamente diverso da 0).

L'ipotesi nulla viene respinta se il valore del parametro stimato p è tale che:

$$|p| > z_{\frac{\alpha}{2}} \cdot \frac{SE}{\text{Errore Standard}}$$

Per $\alpha = 0,05$, $z_{\frac{\alpha}{2}} = 1,96$, avremo:

$$\frac{|p|}{SE} > 1,96$$

Il rapporto $\frac{|p|}{SE}$ viene definito *valore - t* del parametro p , dunque, se:

$$|\text{valore} - t| > 1,96$$

verrà respinta l'ipotesi $\mathbf{H}_0: \mathbf{P} = \mathbf{0}$.

In generale, verranno esclusi dal modello tutti i parametri per cui $|\text{valore} - t| < 2$. L'esclusione dovrà essere fatta un parametro alla volta, con successiva nuova stima del modello, in quanto l'eliminazione anche solo di un parametro modifica tutti gli altri.

4.10.2 Inclusione di nuovi parametri (*indici di modifica*).

Gli *indici di modifica* consentono di includere nel modello dei parametri significativi inizialmente non previsti.

Per ogni parametro *fisso*, cioè non incluso nel modello, viene calcolato di quanto diminuirebbe la *statistica T* del chi-quadrato se esso venisse «liberato», cioè incluso nell'analisi, eliminandone l'uguaglianza a 0. L'indice di modifica è approssimativamente un chi-quadrato con un grado di libertà. Esso è quindi significativo se il suo valore è superiore a 4.

La scelta dei parametri da liberare avviene sia su una selezione preliminare da parte del ricercatore, sia sulla base di un criterio puramente matematico: si individuano i parametri con gli indici di modifica più elevati e si introducono, uno alla volta, nell'analisi, stimando nuovamente il modello.

4.11 Stima PLS

Nel caso in cui ci disponiamo di molte variabili e poche osservazioni la regressione PLS rappresenta una delle tecniche più appropriate.

Il vantaggio di questo metodo è quello di essere in grado di mettere in relazione un blocco l'insieme delle variabili della matrice (X) con una o più variabili di risposte della matrice Y risposte. La metodologia PLS rappresenta un metodo di regressione piuttosto robusto, nel senso che fornisce buoni risultati anche quando i predittori siano correlati tra loro ed è efficace anche quando il numero delle variabili è superiore al numero degli osservazioni. La fig 2 è una rappresentazione schematica del metodo PLS:

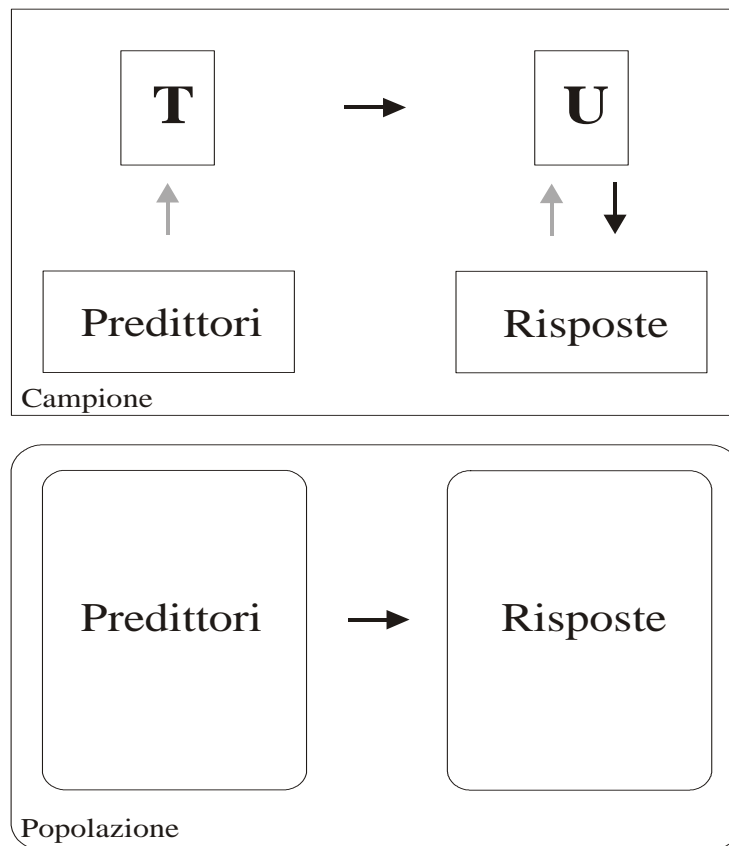


Fig 2 Metodo Pls

.L'obiettivo è utilizzare i predittori per spiegare le variabili di risposta. Ciò è raggiunto indirettamente estraendo le variabili latenti T ed U rispettivamente dalle variabili indipendenti e dipendenti misurate su un campione. I fattori T, estratti da X, sono utilizzati per spiegare i fattori U, estratti da Y, che successivamente sono utilizzati per predire le variabili di risposta.

Da un punto di vista computazionale, dopo aver centrato e standardizzato, X ed Y vengono scomposte contemporaneamente nel seguente modo:

$$X = TP + E \quad (T, E \text{ sono } n \times k, P \text{ è } k \times p)$$

$$Y = UQ + F^* \quad (U, F^* \text{ sono } n \times k, Q \text{ è } k \times q)$$

dove

$k < p$ è il numero di vettori associati ad X.

E è la matrice dei residui di X al k-esimo stadio; quando $k = q$, $E = 0$

F^* è uno step intermedio per ottenere la matrice dei residui di Y al k-esimo stadio

T, U sono le variabili latenti rispettivamente di X ed Y.

La relazione tra X ed Y è modellata attraverso la seguente operazione:

$$U = TB \quad (B \text{ è } k \times k)$$

che produce:

$$Y = TBQ + F$$

dove F è la matrice dei residui di Y al k-esimo stadio

Il metodo PLS è un metodo iterativo ove ogni componente principale (chiamata in PLS "variabile latente") viene determinata per tener conto contemporaneamente sia della massima varianza dei predittori sia della massima correlazione di ogni componente con la variabile di risposta.

In termini più tecnici il criterio del PLS è quello di ricercare le componenti t_h (combinazioni lineari di X) e q_h (combinazioni lineari delle Y) tale che la covarianza tra esse sia massima.

Poichè

$$Cov(t_h, q_h) = r \sqrt{Var(t_h)Var(q_h)}$$

dove r rappresenta il coefficiente di correlazione tra t_h e q_h , si evince che per massimizzare tale covarianza, si devono massimizzare simultaneamente: $Var(t_h)$, $Var(q_h)$, $r(t_h, q_h)$

Le operazioni sopra descritte vengono effettuate mediante l'algoritmo noto in letteratura come PLS2 (PLS1 è la versione utilizzata quando c'è una sola variabile di risposta), descritto in Wold et al. (1983), Martens e Naes (1985).

CAPITOLO QUINTO

ENTROPIA

Sommario: **5.1** L'Entropia di Shannon come misura della Casualità - **5.2** La Massima Entropia Generalizzata - **5.3** I modelli ad equazioni strutturali di Massima Entropia Generalizzata - **5.4** L'Algoritmo GME per SEM **5.5** Simulazione

5.1 L'Entropia di Shannon come misura della Casualità

L'approccio della massima entropia si basa sulla funzione di Entropia di Shannon (1948)⁵⁶, il quale definì un metodo assiomatico per misurare l'incertezza (stato di conoscenza) di una collezione di eventi. Considerando una variabile causale discreta X di eventi x_i con $i=1, \dots, N$ come sorgente di informazione, con probabilità associata p_i , la quale consente di calcolare indici statistici quali la media o la varianza, ma non permette di stimare in termini probabilistici la casualità associata al fenomeno nella sua globalità. Shannon definisce l'entropia dell'informazione (o stato della conoscenza) mediante un metodo assiomatico basata su una funzione di probabilità unica:

$$H(P) = -k \sum_{i=1}^k p_i \ln(p_i); \quad (1)$$

⁵⁶ Shannon C.E. (1948). A mathematical Theory of Communications, *Bell System Technical Journal*, 27, 379-423

Dove k è solitamente una costanza pari ad 1, $0\ln(0)=0$, e $\{ \sum_i p_i = 1 \}$ l'ammontare delle $(-\ln(p_i))$ è chiamato l'ammontare delle informazioni proprie in merito all'evento i (self information). La media delle informazioni proprie è definito entropia. La migliore approssimazione è scegliere p_i , in modo da massimizzare la precedente funzione rispetto i vincoli di Consistenza e Normalizzazione.

Un analogo concetto era già conosciuto in termodinamica ed in meccanica statistica⁵⁷.

Golan et al (1996) sviluppa la procedura GME per i *general linear econometric* model, per la stima mediante il Principio della Massima Entropia, quando i dati sono incompleti in presenza degli outlier , o fortemente correlati (ill posed problems). Al-Nasser et al (2000) applica il metodo GME per la stima dei parametri dei modelli ad equazioni strutturali (MES), i cui risultati sono confrontati con altri metodi di stima consolidati in letteratura nell'ambito della Customer Satisfaction mediante simulazioni ed applicazioni.⁵⁸

Jaynes (1957, 1968)⁵⁹ fu il primo ad usare l'Entropia di Shannon come strumento per misurare la probabilità di una distribuzione , sottoforma di valore atteso, in presenza di determinati vincoli generati dai dati definendo il principio di massima entropia (MEP). Il MEP può essere formalizzato

⁵⁷ Clausius e Boltzmann diedero la prima espressione funzionale per l'entropia come misura del grado di disordine esistente in un sistema termodinamico

⁵⁸ Ciavolino E. Al Nasser A.D., D'Ambra A. (2006). *The Generalized Maximum Entropy Estimation method for the Structural Equation Model*, GFKL 2006, Berlino Marzo.

⁵⁹ Jaynes E.T. (1957). Information Theory and Statistical Mechanics, *The Physical Review* 106 (4), 620-630, May 15, 1957.

Jaynes E.T. (1968). Prior Probabilities, *IEEE Transactions On Systems Science and Cybernetics*, vol. sec-4, no. 3, 227-241

sulla base della variabile casuale X precedentemente definita, di cui però si vuole stimare la probabilità. Si definiscono le funzioni $\{f_1(x_1), f_2(x_2), \dots, f_T(x_1)\}$, le quali rappresentano le informazioni disponibili sui dati, ed i valori misurabili $\{y_1, y_2, \dots, y_T\}$, ottenuti sui dati training, oppure ottenuti mediante informazioni a priori sul fenomeno oggetto di studi⁶⁰.

Il MEP stima la distribuzione di probabilità di Shannon (1) sulla base di vincoli di coerenza (2), espressi in forma di valori attesi, e vincoli di normalizzazione, cioè $\{ \sum_i p_i = 1 \}$:

$$\{ \sum_i^n f_i(x_i) p_i = y_t \} \text{ con } t= 1, 2, \dots, T \quad (2)$$

Tale problema può essere risolto utilizzando il metodo dei moltiplicatori di Lagrange (Jaynes 1957)⁶¹.

5.2 La Massima Entropia Generalizzata

Gola et al. (1996)⁶² come accennato, hanno proposto un metodo alternativo per la stima dei parametri per i modelli di regressione in caso di problemi di mal-condizionati (ill-posed problem), come estensione della misura di entropia di Shannon e generalizzazione del MEP di Jaynes. Il metodo è chiamato *Generalized Maximum Entropy* (GME) e si basa sulla

⁶⁰ Nell'esperimento del lancio del dado di Jaynes, la y , per un dado non truccato, è pari a 3,5.

⁶¹ Purtroppo il risultato è un problema in forma non chiusa, che viene risolto attraverso l'utilizzo di opportune tecniche numeriche.

⁶² Golan A., Judge G. & Karp L. (1996). A maximum entropy approach to estimation and inference in dynamic models of counting fish in the sea using maximum entropy, *Journal of Economic Dynamic and Control*, 20, 559-582

ri-parametrizzazione e ri-formulazione di un modello lineare generalizzato

$y = X\beta + \varepsilon$; con n unità ed m variabili, in modo da stimare i parametri

all'interno della formulazione del MEP, secondo la seguente equazione:

$$y_{n,1} = X_{n,m} \beta_{m,1} + \varepsilon_{n,1} = X_{n,m} Z_{m,mM} p_{mM,1} + V_{n,nN} w_{nN,1} \quad (3)$$

Le matrici Z e V sono diagonali ed il generico elemento è rappresentato rispettivamente dai vettori

$$z_k = [-c; -c/2; 0; c/2, c] \text{ e } v_k = [-c; -c/2; 0; c/2, c].$$

Tali vettori definiscono le variabili di supporto, chiamati fixed points, con 5 elementi ($M=N=5$), equamente distribuiti in modo simmetrico rispetto a zero. I vettori p e w associati, sono le probabilità da stimare mediante massimizzazione della funzione di entropia di Shannon $H(P,W) = -p \ln p - w \ln w$ soggetta a vincoli di normalizzazione e di coerenza. La formulazione matriciale è di seguito riportata, dove $k = mM$ e $T = nN$:

$$\beta = Zp = \begin{bmatrix} z_1 & 0 & \cdots & 0 \\ 0 & z_2 & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & z_k \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{bmatrix};$$

$$\varepsilon = Vw = \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & v_k \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix}$$

Gli steps per il calcolo dell'algoritmo GME sono 4:

Step1

Riparametrizzare i parametri incogniti ed i termini di errore, come una combinazione convessa di valori attesi di una variabile casuale discreta;

Step2

Riscrivere il modello usando i parametri e gli errori ri-parametrizzati (tale formulazione rappresenta il vincolo sui dati);

Step3

Formulare il problema GME come un problema di programmazione non lineare secondo la seguente formulazione:

Funzione Obiettivo = Shannon's Entropy Function

1. Vincoli di consistenza, cioè la nuova formulazione del modello;
2. Vincoli di normalizzazione.

Step4

Risolvere il problema di programmazione non-lineare usando un metodo di calcolo numerico.

5.3 I modelli ad equazioni strutturali di Massima Entropia Generalizzata

I MES basati sulla stima GEM, partono dalla classica formulazione di Jöreskog (1970)⁶³; in cui si distinguono l'equazione del modello strutturale (4) che rappresenta il legame lineare tra le variabili latenti, e le equazioni del modello di misurazione, di tipo endogeno (5) ed esogeno (6) (Outer

⁶³ Jöreskog K.G. (1970). *A general method for estimating a linear Structural Equation System*, in Goldberger e Duncan, 85-112

model), e definiscono le relazioni tra le variabili latenti e manifeste. (fig.1 esempio di mod strutturale)

$$\eta_{(mx1)} = B_{(mxm)} \eta_{(mx1)} + \Gamma_{(mxn)} * \xi_{(nx1)} + \tau_{(mx1)} \quad (4)$$

$$y_{(px1)} = \Lambda^y_{(pxm)} \eta_{(mx1)} + \varepsilon_{(px1)} \quad (5)$$

$$x_{(qx1)} = \Lambda^x_{(qxn)} \xi_{(nx1)} + v_{(qx1)} \quad (6)$$

Fig.1 esempio di mod strutturale

Il modello **Strutturale**

$$\eta_1 = \gamma_{1,1} \xi_1 + \gamma_{1,2} \xi_2 + \tau_1$$

Il modello di Misurazione **Endogeno**

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} \lambda_{1,1}^y \\ \lambda_{2,1}^y \\ \lambda_{3,1}^y \\ \lambda_{4,1}^y \\ \lambda_{5,1}^y \end{bmatrix} \cdot \eta_1 + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}$$

Il modello di Misurazione **Esogeno**

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} \lambda_{1,1}^x & 0 \\ \lambda_{2,1}^x & 0 \\ \lambda_{3,1}^x & 0 \\ 0 & \lambda_{4,2}^x \\ 0 & \lambda_{5,2}^x \end{bmatrix} \cdot \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{bmatrix}$$

Sul modello evidenziato dobbiamo fare delle ipotesi,

- le variabili sono supposte siano centrate, quindi:

$$E(\eta) = E(\tau) = 0; E(Y) = E(\varepsilon) = 0; E(X) = E(\delta) = 0; E(\xi) = 0;$$

- le variabili indipendenti e gli errori sono in correlati:

$$\text{nella stessa equazione } E(\xi v') = 0; E(\eta \varepsilon') = 0; E(\xi \delta') = 0$$

$$\text{fra equazioni: } E(\eta \delta') = 0; E(\xi \varepsilon') = 0;$$

- gli errori tra le diverse equazioni sono fra loro incorrelati

$$E(\tau\epsilon') = 0; E(\tau\delta') = 0; E(\epsilon\delta') = 0;$$

- la matrice B non è singolare.

In questo modello si considerano m ed n variabili latenti, rispettivamente endogene ed esogene, con p e q variabili manifeste, endogene ed esogene.

Le tre equazioni possono essere riformulate definendo un solo modello:

$$Y_{(p,1)} = \Lambda_{(p,m)}^y (I_{(m,m)} - B_{(m,m)})^{-1} \{ \Gamma_{(m,n)} \Lambda_{(n,q)}^{x-1} (X_{(q,1)} - \delta_{(q,1)}) + \tau_{(m,1)} \} + \epsilon_{(p,1P)} \quad (7)$$

I parametri $(B, \Gamma, \Lambda^x, \Lambda^y, \delta, \tau, \epsilon)$ dell'equazione (7) sono ri-parametrizzati ⁶⁴

in forma di valori attesi di una variabile casuale discreta. In questo modo è possibile riscrivere il modello (7) mediante la funzione di entropia di Shannon (1). La massimizzazione di tale funzione, sottoposta a vincoli di coerenza e di normalizzazione consente la stima dei parametri incogniti.

Tale modello di stima rappresenta un'ottima alternativa al Partial Least Squares (PLS), soprattutto quando il campione è di piccole dimensioni ed i dati sono fortemente correlati (Ciavolino et al 2006)⁶⁵.

Un ulteriore vantaggio è il calcolo di un indice chiamato *Normalized Entropy Measure* (Golan et al 1996)⁶⁶, che quantifica il livello di informazione generato dal modello sulla base dei dati rilevati, il quale offre una misura globale del livello di bontà del MES. L'indice di entropia normalizzata (Normalized entropy measure) è espresso mediante la seguente formulazione:

⁶⁴ Al-Nasser A.D. (2003). Customer Satisfaction Measurement Models: *Generalized Maximum Entropy Approach*. Pak Journal of Statistics, 19 (2), 213-226

⁶⁵ Ciavolino E. Al Nasser A.D., D'Ambra A. (2006). *The Generalized Maximum Entropy Estimation method for the Structural Equation Model*, GFKL 2006, Berlino Marzo.

⁶⁶ Golan A., Judge G. & Karp L. (1996). A maximum entropy approach to estimation and inference in dynamic models of counting fish in the sea using maximum entropy, *Journal of Economic Dynamic and Control*, 20, 559-582

$$S(\hat{p}) = \frac{-p' \ln p}{K \ln M} \quad (8)$$

Tale misura indica la riduzione di incertezza dell'informazione, dove k è il numero di predittori ed M il numero di fixed points. La quantità $K \ln M$ rappresenta la massima incertezza, così se $S(\hat{p})$ assume valore 0, significa nessuna incertezza, se assume valore 1, significa massima incertezza.

5.4 L'Algoritmo GME per SEM

Step1: riparametrizzare il modello

Il modello (4) (5) (6), precedentemente formalizzato può essere riscritto come:

$$Y_{(p*1)} = \Lambda_{(p*m)}^y (I_{(m*m)} - B_{(m*m)})^{-1} \{ \Gamma_{(m*n)} \Lambda_{(n*q)}^{x-1} (X_{(q*1)} - \delta_{(q*1)}) + \tau_{(m*1)} \} + \varepsilon_{(p*1)} \quad (9)$$

Riparametrizzare i parametri incogniti ed i termini di errore, come una combinazione convessa di valori attesi di una variabile casuale discreta con due o più set di punti:

$$\beta_{jk} = \sum_{s=1}^S z_{jks} b_{jks}, \sum_{s=1}^S b_{jks} = 1, j = 1, 2, \dots, m; k = 1, 2, \dots, m$$

$$\gamma_{ij} = \sum_{l=1}^L g_{ijl} f_{ijl}, \sum_{l=1}^L f_{ijl} = 1, j = 1, 2, \dots, m; i = 1, 2, \dots, n$$

$$\lambda_{qi}^x = \sum_{a=1}^A L_{qia}^x d_{qia}^x, \sum_{a=1}^A d_{qia}^x = 1, q = 1, 2, \dots, q; i = 1, 2, \dots, n$$

$$\lambda_{pj}^y = \sum_{c=1}^C L_{pjc}^y d_{pjc}^y, \sum_{c=1}^C d_{pjc}^y = 1, p = 1, 2, \dots, p; j = 1, 2, \dots, m$$

$$\xi_j = \sum_{t=1}^T v_{jt} w_{jt}, \sum_{t=1}^T w_{jt} = 1, j = 1, 2, \dots, m$$

$$\delta_q = \sum_{r=1}^R v_{qr}^x w_{qr}^x, \sum_{r=1}^R w_{qr}^x = 1; q = 1, 2, \dots, q$$

$$\varepsilon_p = \sum_{e=1}^E v_{pe}^y w_{pe}^y, \sum_{e=1}^E w_{pe}^y = 1; p = 1, 2, \dots, p$$

GME calcola le probabilità incognite, che rappresentano la funzione di distribuzione di una variabile casuale.

I parametri e gli errori incogniti non sono in forma di probabilità e la loro somma non è pari ad uno. Quindi i parametri devono essere riscritti in termini di probabilità, cioè bisogna riparametrizzare i termini incogniti come valori attesi di una variabile causale discreta

Step2: Riscrivere il modello

Usando questa formula di ri-parametrizzazione il modello (9) può essere riscritto come:

$$y_p = \psi(b, f, d^x, d^y, w^x, w^y, w)$$

Dove ψ è ottenuta sostituendo i termini dell'equazione precedente in forma di probabilità:

$$\begin{aligned} \psi(b, f, d^x, d^y, w^x, w^y, w) = & \left\{ \left(\sum_j \sum_c L_{pj}^y d_{pj}^y \right) \left(1 - \sum_j \sum_k \sum_s z_{iks} b_{iks} \right) \right\} \bullet \\ & \left\{ \left(\sum_q \sum_i \sum_a L_{qia}^x d_{qia}^x \right) \left(\sum_i \sum_j \sum_l g_{ijl} f_{ijl} \right) \left(\sum_q \left[x_q - \sum_r v_{qr}^x w_{qr}^x \right] \right) + \left(\sum_j \sum_t v_{jt} w_{jt} \right) \right\} + \sum_e v_{pe}^y w_{pe}^y \end{aligned}$$

(10)

(v^x, v^y, v) sono scelti in maniera tale da essere simmetrici intorno allo zero

per ogni $j, q, e p$.

Step3 Riformulazione

Il sistema GME può essere espresso come un problema di programmazione non-lineare, soggetto a vincoli lineari. Bisogna massimizzare la funzione obiettivo rispettando i vincoli di (1) consistenza e (2) normalizzazione.

Il sistema è di seguito formalizzato:

$$\max H(b, f, d^x, d^y, w^x, w^y, w)$$

(1)

$$y_p = \psi(b, f, d^x, d^y, w^x, w^y, w)$$

(2)

$$\sum_{s=1}^S b_{jks} = 1; j = 1, 2, \dots, m; k = 1, 2, \dots, m$$

$$\sum_{l=1}^L f_{ijl} = 1; j = 1, 2, \dots, m; i = 1, 2, \dots, m$$

$$\sum_{a=1}^A d_{qia}^x = 1, q = 1, 2, \dots, q; i = 1, 2, \dots, n$$

$$\sum_{c=1}^C d_{pjc}^y = 1, p = 1, 2, \dots, p; j = 1, 2, \dots, m$$

$$\sum_{t=1}^T w_{jt} = 1, j = 1, 2, \dots, m$$

$$\sum_{r=1}^R w_{qr}^x = 1, q = 1, 2, \dots, q$$

$$\sum_{e=1}^E w_{pe}^y = 1, p = 1, 2, \dots, p$$

Avremo che:

$$\begin{aligned}
H(b, f, d^x, d^y, w^x, w^y, w) = & \\
& - \sum_j \sum_k \sum_s b_{jks} \ln(b_{jks}) - \sum_i \sum_j \sum_l f_{ijl} \ln(f_{ijl}) - \sum_q \sum_i \sum_a d_{qia}^x \ln(d_{qia}^x) \\
& - \sum_p \sum_j \sum_c d_{pjc}^y \ln(d_{pjc}^y) - \sum_j \sum_t w_{jt} \ln(w_{jt}) - \sum_q \sum_r w_{qr}^x \ln(w_{qr}^x) \\
& - \sum_p \sum_c w_{pc}^y \ln(w_{pc}^y)
\end{aligned}$$

Step4 Soluzione

Tuttavia per risolvere questo sistema di programmazione non lineare dovrebbe essere usato un metodo di calcolo numerico & Lagrangiana

5.5 Simulazione

Per illustrare il metodo di stima GME abbiamo condotto uno studio di simulazione usando un modello semplificato che è stato usato per costruire l'indice di CS svedese, proposto da *Claes C. et al (1999)*.

Tale modello è costituito da tre variabili esogene $\xi_1 \xi_2 \xi_3$; ed una variabile endogena η . La struttura è così definita :

$$\eta = \gamma_1 \xi_1 + \gamma_2 \xi_2 + \gamma_3 \xi_3 + \zeta$$

Dove $\gamma_1 \gamma_2 \gamma_3$ sono i coefficienti di regressione e ζ è l'errore standard. Le variabili manifeste sono espresse come x per le variabili ξ e y per la variabile η .

I modelli per le variabili ξ sono espressi come segue (*Bragozzi and Fornell (1982)*):

$$\xi_1 = \pi_1 x_1 + \pi_2 x_2 + \pi_3 x_3 + \delta_1$$

$$\xi_2 = \pi_4 x_4 + \pi_5 x_5 + \pi_6 x_6 + \delta_2$$

$$\xi_3 = \pi_7 x_7 + \pi_8 x_8 + \pi_9 x_9 + \delta_3$$

Dove η sono i coefficienti di regressione.

Il modello di misura per la variabile manifeste η invece:

$$y_1 = \lambda_1 \eta + \varepsilon$$

$$y_2 = \lambda_2 \eta + \varepsilon_2$$

$$y_3 = \lambda_3 \eta + \varepsilon$$

$$y_4 = \lambda_4 \eta + \varepsilon$$

Dove λ sono i coefficienti

Considerando questo modello strutturale la simulazione è stata fatta

sotto queste condizioni:

- Considerando 100 campioni casuali con $N = 15; 20; 25; 30; 35; 40$.
- per il modello formativo i valori delle x sono stati generati da una distribuzione simmetrica Beta con i parametri (6,6)
- tutti i coefficienti π assumono valore 1/3;
- I coefficienti γ sono inizializzati con (0.8, 0.1, 0.1)
- I coefficienti λ sono inizializzati con (1.1, 1.0, 0.9, 0.8)
- Gli errori δ e di ε sono generati da una distribuzione Uniforme $U(0,1)$; mentre ξ sono generati da una distribuzione Normale Standardizzata;

Sotto queste condizioni ed effettuando una simulazione con il

FORTRAN, abbiamo ottenuto i seguenti risultati:

N	MSE ($\bar{\hat{\pi}}$)	MSE ($\hat{\gamma}_1$)	MSE ($\hat{\gamma}_2$)	MSE ($\hat{\gamma}_3$)	MSE($\bar{\hat{\lambda}}$)
15	7,406E-3	4,266E-2	6,679E-4	6,675E-4	7,406E-2
20	4,788E-3	2,081E-2	5,493E-4	4,970E-4	4,788E-2
25	4,046E-3	2,030E-2	5,111E-4	3,449E-4	4,606E-2
30	3,974E-3	1,965E-2	4,042E-4	2,577E-4	3,009E-2
40	3,915E-3	8,032E-2	3,827E-4	1,348E-4	1,470E-2

Tab1 MES con stima dei parametri GME

N	MSE ($\bar{\hat{\pi}}$)	MSE ($\hat{\gamma}_1$)	MSE ($\hat{\gamma}_2$)	MSE ($\hat{\gamma}_3$)	MSE($\bar{\hat{\lambda}}$)
15	2,716E-1	6,456E-1	1,474E-1	1,570E-1	2,628
20	2,037E-1	4,842E-1	1,105E-1	1,178E-1	1,971
25	1,629E-1	3,874E-1	8,845E-2	9,425E-2	1,57
30	1,086E-1	3,228E-1	7,370E-2	7,854E-2	1,314
40	0,148E-1	2,421E-1	5,528E-2	5,890E-2	9,857

Tab2 MES con stima dei parametri PLS

Dove ($\bar{\hat{\pi}}$) è la media delle stime dei coefficienti delle variabili ξ esogene

($\bar{\hat{\lambda}}$) è la media delle stime dei coefficienti delle variabili endogene η .

Da questi risultati possiamo notare che la metodologia GME risulta essere migliore rispetto al PLS e fornisce una migliore stima soprattutto con N piccolo, in effetti presenta il valore dell'errore standard più piccolo.

CAPITOLO SESTO

CASO STUDIO

Sommario: **6.1** Lo Studio della CS nel Punto Vendita - **6.2** Modello di valutazione della **6.3** Confronto PLS – MSE **6.4** Matrice Degli Interventi **6.5** Supporto Alle Decisioni – **6.6** Criteri di Simulazione – **6.7** Confronto MSE variabili Latenti: Caso Generale – **6.7.1** Confronto MSE variabili Manifeste: Caso Generale – **6.8** Confronto MSE variabili Latenti: Missing Value – **6.8.1** Confronto MSE variabili Manifeste: Missing Value – **6.9** Confronto MSE variabili Latenti: Outliers – **6.9.1** Confronto MSE variabili Manifeste: Outliers – **6.10** Confronto MSE variabili Latenti: Multicollinearità 0,3 (bassa) - **6.10.1** Confronto MSE variabili Manifeste: Multicollinearità 0,3 (bassa) – **6.11** Confronto MSE variabili Latenti: Multicollinearità 0,9 (alta) – **6.11.1** Confronto MSE variabili Manifeste: Multicollinearità 0,9 (alta) – **6.12** Analisi dei risultati

6.1 Lo Studio della CS nel Punto Vendita

L'interesse nei confronti delle problematiche inerenti la qualità e cortesia offerta nei supermercati - ipermercati, intesa come capacità di soddisfare i bisogni impliciti ed espliciti della clientela, a causa di una presenza molto accentuata della concorrenza e dalla voglia di “conquistare” sempre più nuove quote di mercato, è diventato molto elevato negli ultimi anni.

Scopo della valutazione della soddisfazione dei clienti e della qualità offerta dal punto vendita (pdv) è quello di riuscire a fidelizzare sempre più il cliente, cercando di cogliere quanti più aspetti possibili, circa le loro abitudini, la loro dieta, la loro fiducia verso il proprio marchio, ma anche di

capire se la struttura è caratterizzata da un adeguato lay-out, se il personale è gentile e disponibile se si ha la necessità di introdurre nuove “referenze”.

Il ns lavoro attuato grazie ad una collaborazione università – responsabili della qualità di una delle più grandi catene di distribuzione presente sul ns territorio, ha come obiettivo quello di raccogliere informazioni principalmente, ma non esclusivamente, su queste categorie di servizi. Del resto i servizi che i pdv offrono, rivestono un ruolo molto importante nel sistema economico dell’azienda, sia in termini di risorse impegnate e quindi di costi di personale da sostenere, ma soprattutto di soddisfazione creata.

L’intero processo di gestione del punto vendita è stato da noi, in concomitanza con il management, suddiviso in fasi attraverso un’aggregazione di attività che concorrono allo stesso obiettivo (Fig 1.)

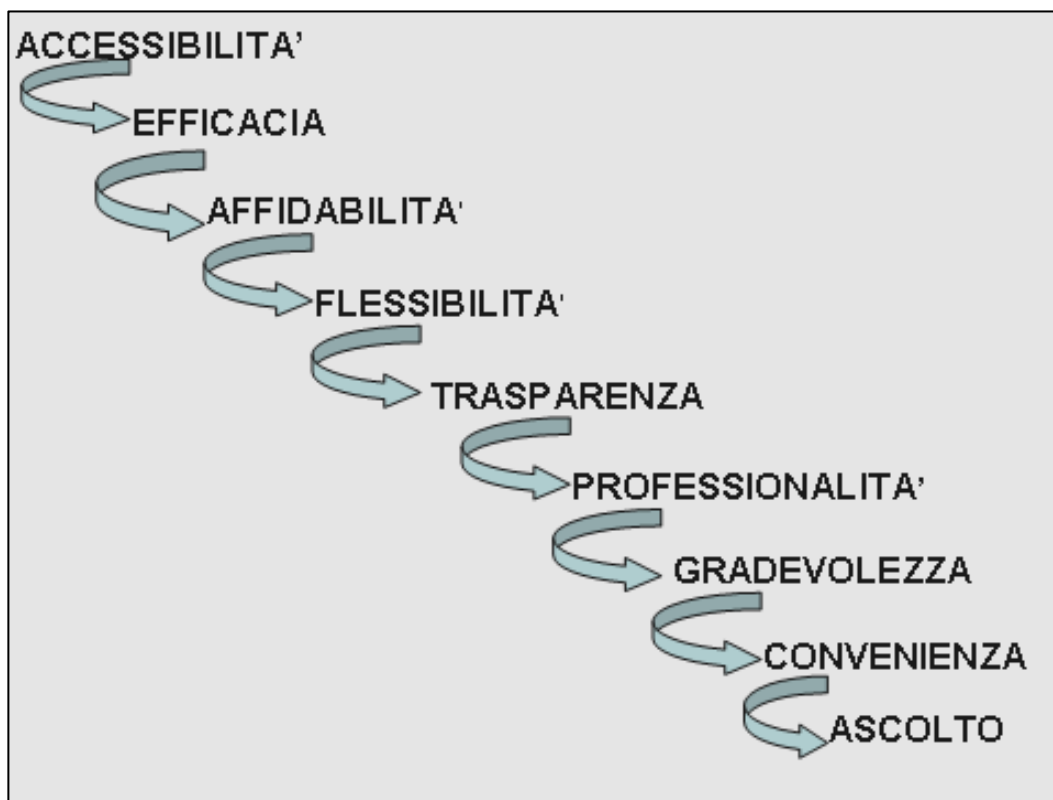


Fig1 Dimensioni latenti della Customer Satisfaction nell'ambito della distribuzione

Dalla scomposizione di tale processo è stata immediata l'individuazione delle variabili su cui indirizzare la qualità del servizio e su cui, di conseguenza valutare la soddisfazione del cliente.

FASI & VARIABILI	OBIETTIVO
1. Accessibilità	Garantire l'accesso al servizio dell'utente in termini di segnaletica stradale e area parcheggio
2. Efficacia	Capacità di rispondere alle necessità dei clienti
3. Affidabilità	Garanzia di continuità del servizio e di Sicurezza nel suo utilizzo

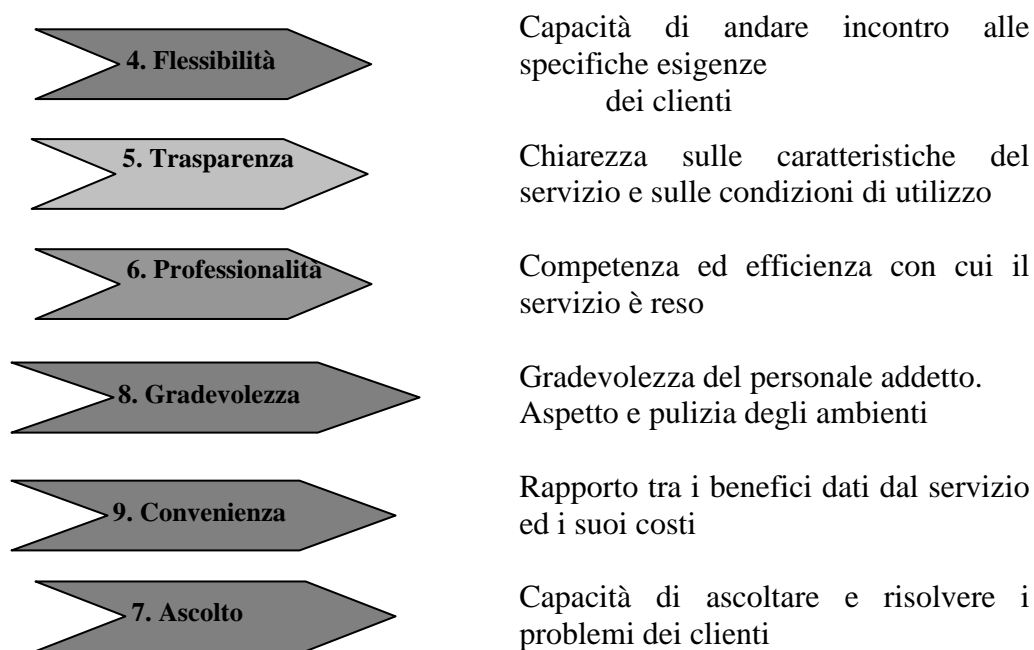


Fig 2 Relazione tra Fasi del Processo ed Obiettivi

Le fasi del processo, identificate dalle 9 caratteristiche della qualità riportate nella fig.2, sono rilevate mediante un questionario di 22 domande specificate nel prossimo paragrafo.

Le caratteristiche della qualità e le possibili interazioni che intercorrono tra di loro, sono state formalizzate nel Modello ad Equazioni Strutturali –MES– (vedi fig3) rispettivamente dai costrutti teorici riportati negli ovali e dalle frecce orientate. Il MES fornisce rappresentazione semplificata dei processi reali, delle cause che agiscono su una variabile dipendente e delle connessioni esistenti tra le diverse cause. I coefficienti del modello sono stati stimati con il metodo GME e sono riportati nel prossimo paragrafo.

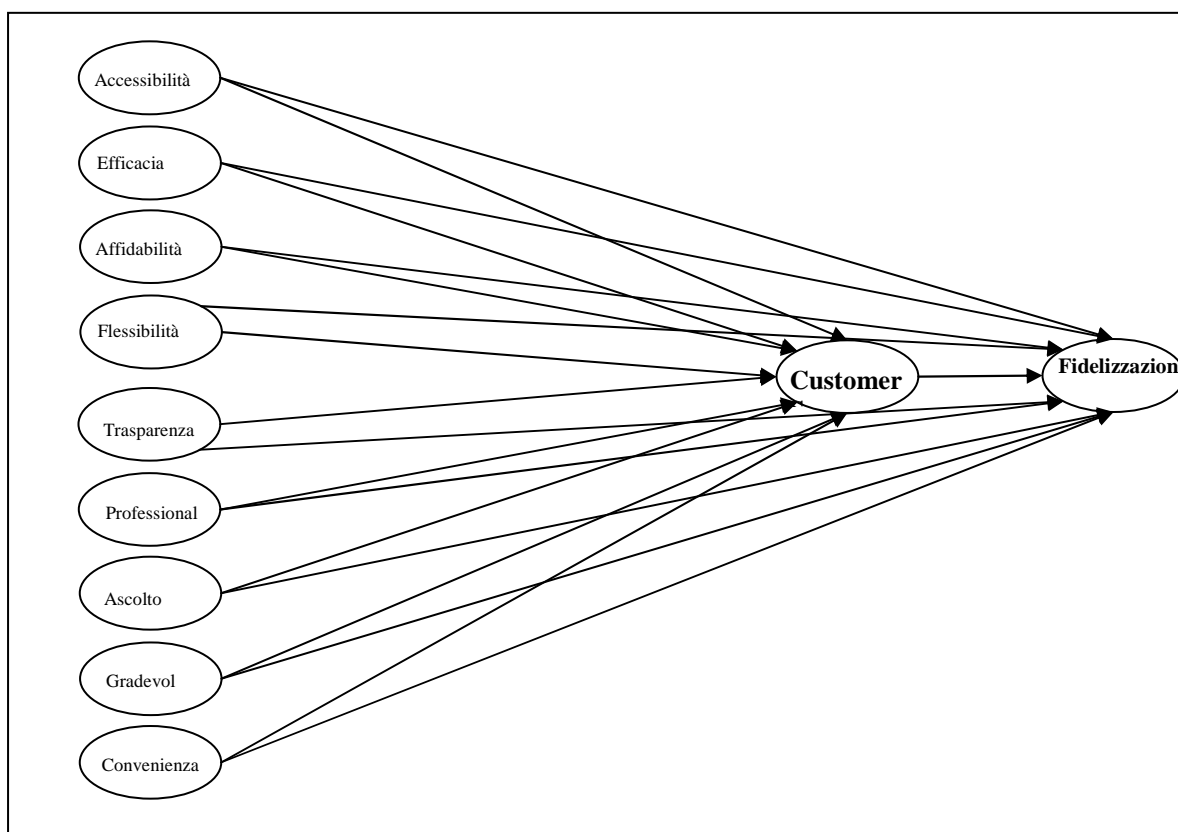


Fig 3 Modello ad Equazioni Strutturali

La rilevazione è avvenuta mediante la somministrazione di questionari ad un campione di 250 clienti.

Il campionamento adottato in questa sede si basa su due diverse unità di rilevazione:

- Clienti Fidelizzati (possessori delle fidelity). Sono i clienti che quotidianamente entrano a fare la spesa nel ns supermercato, si prestano ad un campionamento di tipo probabilistico, perché in questo caso il responsabile del Pdv possiede la lista completa dei clienti possessori delle fidelity e quindi può associare ad ogni singolo cliente la probabilità di entrare a far parte del campione. Il disegno di

campionamento può essere, stratificato per sesso, o per fasce di età, sulla base della lista dei clienti etc.

- Clienti occasionali. Sono quei clienti che non entrano quotidianamente nella ns struttura e pertanto sono privi di fidelity card. Il tipo di campionamento adottato è un campionamento casuale a grappoli (cluster sampling). Il campione viene formato selezionando casualmente delle fasce orarie di maggiore densità (tarda mattinata e tardo pomeriggio) ed in determinati giorni della settimana (sabato e martedì) a partire dai quali si può procedere ad una ulteriore selezione campionaria di unità statistiche.

2 Modello di valutazione della CS

Il modello da noi costruito è stato strutturato in maniera tale da sottoporre ai clienti 22 domande le quali rappresentano le 9 caratteristiche della qualità.

Var. Latenti	VARIABILI MANIFESTE	ID
Accessibilità	Presenza segnaletica stradale	P1
	Disponibilità parcheggi	P2
Efficacia	Erogazione serv da parte del personale	P3
	Il personale è disposto ad aiutarmi	P4
Affidabilità	Qualità dei prodotti	P5
	Il personale mi ispira fiducia	P6
Flessibilità	Il personale soddisfa ogni esigenza	P7
	Il personale è disponibile	P8
Trasparenza	I prezzi sono ben esposti	P9
	I messaggi pubbl coincidono con le offerte	P10

Professionalità	Il personale risulta essere professionale	P11
	La gestione nel complesso	P12
Ascolto	Il personale mi ascolta	P13
	Il personale risolve ogni problema	P14
Gradevolezza	Gradevolezza del personale	P15
	Pulizia/igiene struttura	P16
Convenienza	Offerte soddisfacenti	P17
	Quantità di merce in offerta disponibile	P18
Customer	Giudizio complessivo sul personale	P19
	Giudizio complessivo sul punto vendita	P20
Fedeltà	Convenienza spesa	P21
	Consiglia il ns punto vendita	P22

Fig 4 Modello Costruito

I questionari sono stati costruiti in modo da comprendere la percezione del contributo delle singole dimensioni da parte di ogni singolo cliente usando una scala ordinale da 1 (valutazione minima) a 6 (valutazione massima), la quale è stata sottoposta ad un procedimento di quantificazione seguendo l'approccio di Thurstone per l'arricchimento delle proprietà aritmetiche della scala.

I coefficienti del modello sono stati stimati mediante il modello GME ed i risultati sono riportati nella fig 5, in cui i valori all'interno delle parentesi rappresentano i valori della statistica T-Student, calcolati mediante la tecnica di ricampionamento bootstrap (vedi appendice) considerando 1000 campioni di dimensione 250.

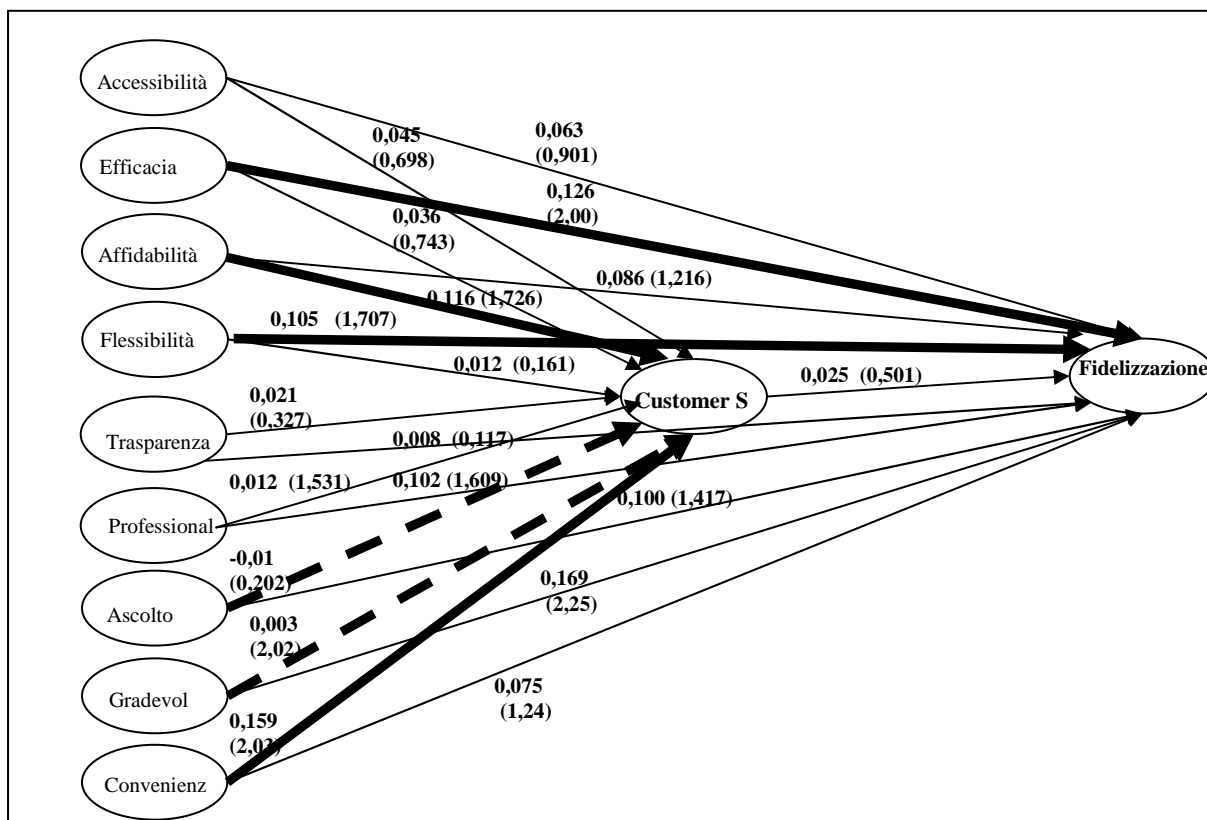


Figura 5 – Il MES e i parametri stimati mediante la GME

Passando alle valutazioni sulle stime dei coefficienti strutturali riportati nella fig. 5, si nota che la maggior parte delle variabili risultano essere significative al livello del 5%, le uniche variabili latenti che non hanno impatto sulla *Customer Satisfaction*, risultano essere la *Gradevolezza*, e l'*Ascolto* (evidenziati da una freccia tratteggiata fig5) oltre al fatto che i clienti lamentano poca disponibilità di parcheggi e più attenzione da parte del personale.

Maggiore contributo al raggiungimento della CS viene dato dall'*Affidabilità* e dalla *Convenienza* (freccia in grassetto fig5) in cui gli elementi di spicco sono la *qualità dei prodotti* e la bontà delle ns offerte.

Al fine di fidelizzare il cliente invece impatto significativo viene dato dalle variabili latenti dell'*Efficacia* e della *Flessibilità*, le quali tengono conto

della disponibilità del personale a rispondere a quelle che sono le singole necessità dei clienti, e di andare incontro a quelle che sono le singole esigenze.

I parametri sono stati stimati anche con i metodi della massima verosimiglianza e del Partial Least Squares. I risultati della prima hanno mostrato una distribuzione non normale nei residui, mentre il PLS (fig.6), pur essendo una buona alternativa in un approccio non parametrico, non consente una misura globale di bontà di adattamento, ricercando dei punti di minimo locale (Wold et Joreskog 1982), inoltre in questa applicazione ha mostrato valori più bassi di Mean Squared Error.

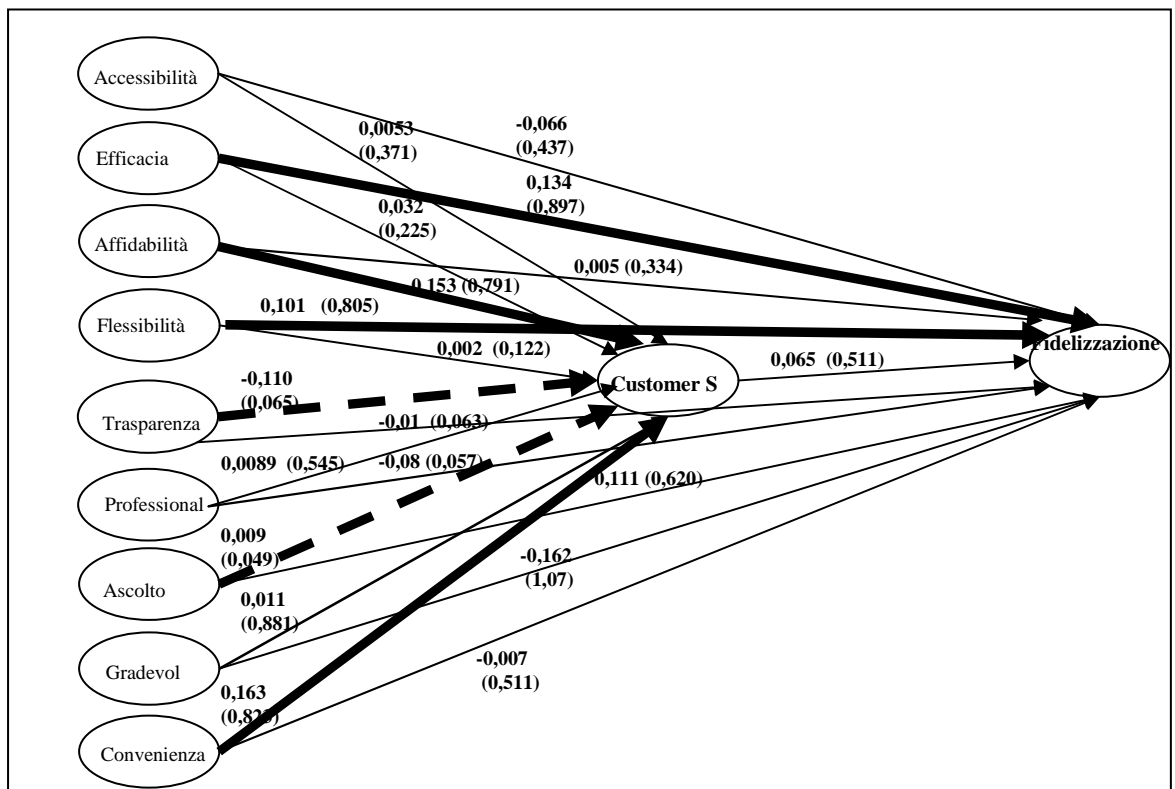


Fig. 6 – Stima Dei parametri mediate PLS

Considerando invece la fig.6 in cui si evidenzia la stima dei coefficienti strutturali con il PLS, le variabili non significative rispetto alla CS,

risultano essere *Ascolto*, e la *Trasparenza* (var evidenziato in tratteggio fig.6) .

Maggiore contributo al raggiungimento della CS viene dato invece dalla *Convenienza* e *Affidabilità* (variabili evidenziate in grassetto fig6), mentre effetto moderato risultano averlo le latenti inerenti la *Gradevolezza*, seguita dalla *Professionalità* e dall'*Accessibilità*.

Mentre per quanto riguarda la fidelizzazione maggiore contributo viene dato dall'*Accessibilità* e dall'*Efficacia*.

6.3 Confronto PLS - GME

Da un confronto dei risultati ottenuti sia con la stima PLS che GME, si è constatato innanzitutto il fatto che l'errore standard delle stime ottenute con GME è sempre più piccolo rispetto a quello ottenute con PLS (vedi fig7), come stato già evidenziato nella simulazione fatta nel capitolo precedente, tutte le stime ottengono path coefficient molto simili, portando essenzialmente alla stessa interpretazione.

	<u>PLS</u>	<u>GME</u>
ACCESSIBILITA' (e.s)	0,053 (0,146)	0,045 (0,0644)
AFFIDABILITA' (e.s.)	0,153 (0,1577)	0,116 (0,0672)
ASCOLTO	0,009	-0,01

(e.s.)	(0,145)	(0,0495)
CONVENIENZA (e.s.)	0,163 (0,182)	0,159 (0,0739)
EFFICACIA (e.s.)	0,032 (0,133)	0,036 (0,0484)
FLESSIBILITA' (e.s.)	0,02 (0,126)	0,012 (0,0745)
GRADEVOLEZZA (e.s.)	0,145 (0,167)	-0,143 (0,0705)
PROFESSIONALITA' (e.s.)	0,089 (0,163)	0,012 (0,0581)
TRASPARENZA	-0,11 (0,171)	0,021 (0,0642)

Fig 7 Confronto PLS – MEG

Anche se come si evince dalla fig.8 la variabilità spiegata con la stima MEG è di gran lunga superiore rispetto alla stima PLS, e ciò non fa altro che confermare la bontà nell'utilizzare suddetta stima, soprattutto in presenza di multicollinearità.

	R²
PLS	7,2%
GME	6,1%

Fig. 8 – Variabilità spiegata da PLS e GME

6.4 Matrice Degli Interventi

I coefficienti strutturali consentono di dividere la variabili esogene in due gruppi, uno ad alto contributo (Ascolto, Accessibilità, Affidabilità, Convenienza, Fedeltà) ed uno a basso contributo (Flessibilità, Trasparenza,

Professionalità, Gradevolezza, Customer) sulla variabile endogena Valutazione Complessiva.

Un'altra importante valutazione viene fuori dallo studio dei valori medi delle variabili latenti, calcolati mediante la ponderazione delle variabili manifeste con i weights e riportati su scala originale, ottenendo in questo modo il livello di gradimento degli utenti relativi ai vari aspetti.

VL	Accessibilità	Efficacia	Affidabilità	Flessibilità	Trasparenza
Media	4,5	3,6	11,6	1,2	2,1
VL	Professionalità	Ascolto	Gradevolezza	Convenienza	
Media	1,2	1,3	14,3	15,9	

Tabella 9 – Valori medi

Anche in questo caso è possibile dividere il gradimento in due gruppi, dove si può leggere un gradimento superiore al valore medio, per gli aspetti che riguardano l'Accessibilità, l'Efficacia, l'Affidabilità, la Convenienza, mentre per la Flessibilità, la Trasparenza, la Professionalità, l'Ascolto e la Gradevolezza il gradimento è sul valore mediano o inferiore.

Considerando entrambe le categorizzazioni suggerite dai coefficienti strutturali e dai valori medi, è possibile costruire una *matrice degli interventi* (tabella 10), la quale, combinando entrambe le informazioni, consente di creare un semplice e valido strumento per il supporto alle decisioni da intraprendere per migliorare il livello di soddisfazione degli utenti.

	<u>SODDISFAZIONE</u>
--	-----------------------------

		<i>Bassa</i>	<i>Alta</i>
<u>IMPORTANZA</u>	<i>Bassa</i>	<i>Flessibilità</i> <i>Professionalità</i> <i>Gradevolezza</i>	
	<i>Alta</i>	<i>Ascolto</i>	<i>Accessibilità</i> <i>Efficacia</i> <i>Affidabilità</i> <i>Convenienza</i>

Tabella 10 – Matrice degli interventi

Il “messaggio” suggerito dalla matrice degli interventi è che i miglioramenti dovrebbero essere approntati per quegli aspetti della qualità per cui l’importanza rispetto alla soddisfazione complessiva è alta e per cui il livello di gradimento è basso. Quindi, dalla tabella 10 è possibile leggere che l’*Ascolto* rappresenta un punto cruciale nel miglioramento della soddisfazione dei clienti del Pdv, mentre il *l’Accessibilità*, *l’Efficacia*, *l’Affidabilità e Convenienza*, rappresentano la punta di eccellenza del servizio offerto.

La *Flessibilità*, *la Professionalità e la Gradevolezza* anche se non raggiungono un’alta valutazione, non risultano influenzare la soddisfazione dei passeggeri, mentre non si aspetti che hanno generato una soddisfazione elevata, nonostante avessero una poca importanza.

Infine, la seguente tabella 11, riporta l’importanza di ogni singola variabile manifesta nel contribuire alla costruzione della variabile latente. La tabella riporta solo le variabili multi-item, in modo da poter individuare quali sono i singoli micro-aspetti che eventualmente risultano più importanti nell’aumento della *Customer Satisfaction*.

Prendendo in considerazione la variabile *Ascolto*, in quanto risulta strategica e necessaria per migliorare il servizio offerto e quindi la soddisfazione, si può notare come l'aspetto più importante sia la *Risoluzione di ogni problema* da parte del personale, seguito dalla disponibilità ad *ascoltare i clienti*.

Var. Latenti	VARIABILI MANIFESTE	ID	Weight	T-Test
Accessibilità	Presenza segnaletica stradale	P1	0,8101	2,4959
	Disponibilità parcheggi	P2	0,4744	1,5683
Efficacia	Erogazione serv da parte del personale	P3	0,8764	3,1449
	Il personale è disposto ad aiutarmi	P4	0,4237	1,4235
Affidabilità	Qualità dei prodotti	P5	0,9632	3,4071
	Il personale mi ispira fiducia	P6	0,2996	0,9722
Flessibilità	Il personale soddisfa ogni esigenza	P7	0,9732	4,1183
	Il personale è disponibile	P8	-0,3287	-1,1205
Trasparenza	I prezzi sono ben esposti	P9	-0,6419	-2,2072
	I messaggi pubbl coincidono con le offerte	P10	0,7388	2,4352
Professionalità	Il personale risulta essere professionale	P11	0,8849	3,2299
	La gestione nel complesso	P12	-0,4922	-1,6846
Ascolto	Il personale mi ascolta	P13	0,3225	1,0032
	Il personale risolve ogni problema	P14	0,9267	3,2933
Gradevolezza	Gradevolezza del personale	P15	0,9774	5,2681
	pulizia/igiene struttura	P16	-0,2045	-0,7705
Convenienza	Offerte soddisfacenti	P17	0,2117	0,7725

	Quantità di merce in offerta disponibile	P18	0,9581	4,3528
Customer	Giudizio complessivo sul personale	P19	0,8887	3,0794
	Giudizio complessivo sul punto vendita	P20	-0,4996	-1,6462
Fedeltà	Convenienza spesa	P21	0,9207	3,3585
	Consiglia il ns punto vendita	P22	0,3413	1,0618

Tabella11 – Weight delle Variabili Manifeste

6.5 Supporto Alle Decisioni

L'analisi critica della valutazione della CS ha consentito la presentazione di un approccio di rilevazione basato su processi e di tecniche statistiche multivariate innovative quali PLS e GME, le cui potenzialità sono evidenziate mediante l'analisi del sistema di ascolto utilizzato dall'azienda oggetto di studi.

I risultati della rilevazione della soddisfazione dei clienti realizzata con un approccio per processi e delle analisi condotte mediante modelli ad equazioni strutturali con stima GME/PLS, ha permesso al Management aziendale del Pdv di conoscere le caratteristiche di qualità su cui il cliente pone particolare attenzione al momento della fruizione del servizio (fare la spesa) e di individuare, nel proprio sistema organizzativo, la fase del processo di gestione su cui effettuare le relative ipotesi di miglioramento (Tabella 10).

Come si può notare dalla seguente tabella, il management aziendale ha posto come priorità di intervento l'*Ascolto*, scegliendo come obiettivo, quello di sensibilizzare sempre più i propri dipendenti nel capire e quindi soddisfare le esigenze dei clienti. Il management ha inoltre scelto di mantenere standard elevati di qualità riguardo l'*Efficacia* l'*Affidabilità* e la *Convenienza*, mentre per l'*Accessibilità* nonostante la soddisfazione dei

clienti sia già elevata si sta valutando la possibilità di trasferire la propria attività in locali a maggiore contributo commerciale sia in termini di maggiore spazio all'interno per una maggiore visibilità dei prezzi e dei prodotti, che all'esterno, per una maggiore area riservata ai parcheggi.

VARIABILI SIGNIFICATIVE	OBIETTIVI
ASCOLTO	<i>Maggiore disponibilità del personale</i>
ACCESSIBILITA'	<i>Maggiore segnaletica Area parcheggi più ampia</i>
EFFICACIA	<i>Migliorare la tempistica ed il modo di erogazione del servizio</i>
AFFIDABILITA'	<i>Avere solo ed esclusivamente prodotti di "marca"</i>
CONVENIENZA	<i>Garantire sempre la presenza dei prodotti in offerta</i>

Tabella 8 – Ipotesi di miglioramento

Un ulteriore sviluppo potrebbe essere la realizzazione di *un sistema integrato di questionari*, che consenta di valutare non solo la soddisfazione dei clienti, ma anche quella del personale dipendente, e dei fornitori, tenendo conto delle interrelazioni tra queste tipologie di stockholders. Inoltre si dovrebbe prevedere anche un *sistema di raccolta dei reclami*, per rilevare il livello di insoddisfazione espressa dai clienti, ottenendo un quadro d'insieme della soddisfazione e insoddisfazione, considerando non solo il cliente principale (il viaggiatore), ma anche il grado di motivazione e consapevolezza delle altre parti interessate ai processi interni.

6.6 Criteri di Simulazione

Il nostro studio di simulazione, è stato condotto considerando:

- 1000 campioni causali di dimensione 15, 30, 50;
- i valori della X sono stati generati da una distribuzione simmetrica $N(2,2)$
- Y è la stima basata sull'equazione riformulata ;
- Gli errori ξ sono generati da una distribuzione Uniforme $U(0,1)$; δ da una distribuzione Normale Standardizzata $N(0,1)$; e ε sono generati da una distribuzione Beta (6,6)
- I coefficienti γ sono inizializzati con (0.8, 0.1, 0.1);
- Valori mancanti: il 10% dei dati generati in X ed Y sono sostituiti in maniera causale secondo la regola della grande media $X(y)$
- Outliers: il 10% dei dati esistenti di X e Y sono stati sostituiti da: $Q1+5*IQR$ dove Q1 è il primo quartile di X(Y), mentre IQR coincide con la zona interquartile
- Multicollinearità: $\xi \sim U(0,1)$; $\delta \sim N(0,1)$; e $\varepsilon \sim \text{Beta}(6,6)$ e tutti i parametri del modello sono inizializzati da 1.

I gradi di associazione tra le variabili esogene sono ottenuti mediante la

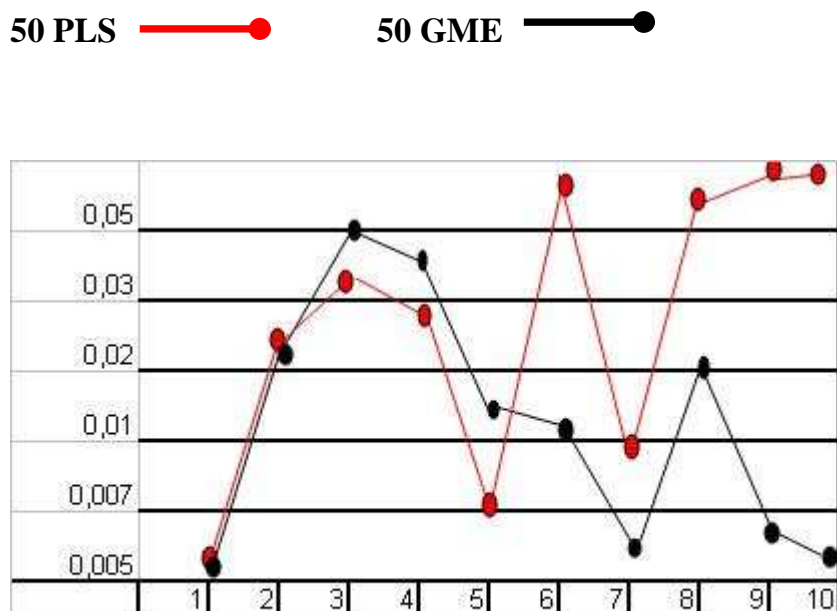
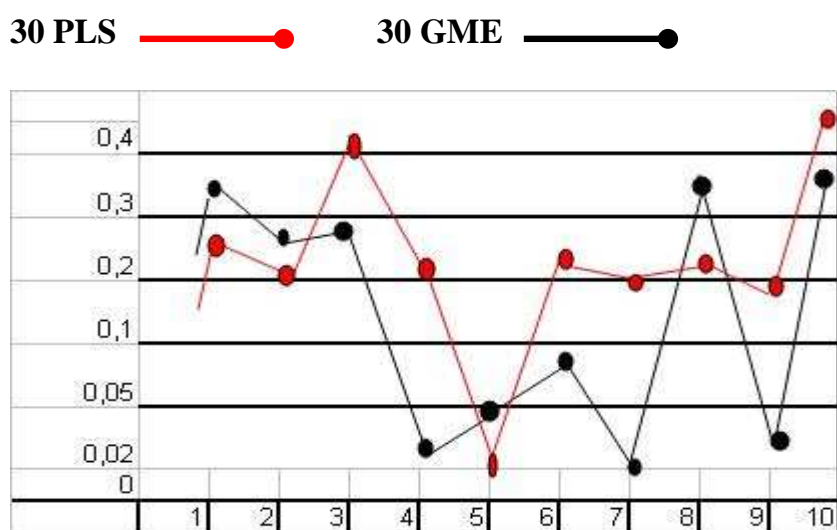
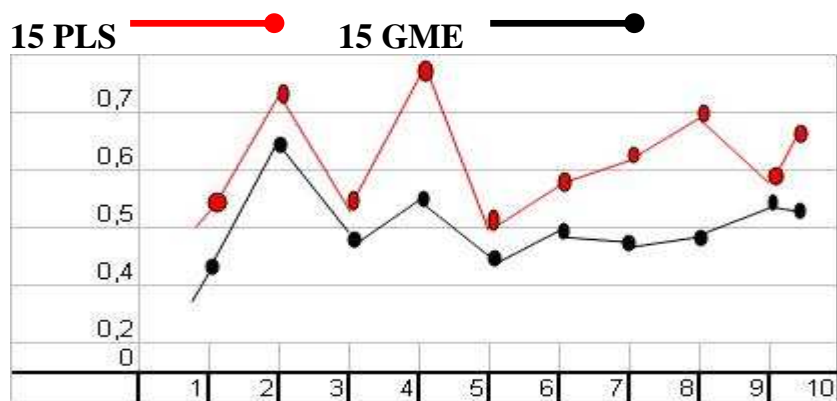
formula: $\zeta_2^{new} = p\zeta_1 + (1 - p^2)\zeta_2^{old}$

Dove il valore tra entrambe le variabili è inizializzato con 0,3 e 0,9.

6.7 Confronto MSE Variabili Latenti: CASO GENERALE

	15		30		50	
	GME	PLS	GME	PLS	GME	PLS
y1	0,4233	0,5234	0,2567	0,1564	0,0041	0,0052
y2	0,6197	0,7145	0,1654	0,1214	0,0219	0,0215
y3	0,4996	0,5123	0,1885	0,3251	0,0125	0,0325
y4	0,5464	0,7335	0,0213	0,1265	0,0199	0,0299
y5	0,4658	0,4996	0,0456	0,0202	0,0154	0,0742
y6	0,5012	0,5779	0,0521	0,1254	0,0145	0,0562
y7	0,4898	0,6002	0,0212	0,1021	0,0079	0,0099
y8	0,4998	0,6998	0,2125	0,1236	0,0214	0,0514
y9	0,5236	0,5335	0,025	0,099	0,0667	0,0864
y10	0,5113	0,6155	0,2324	0,3213	0,0547	0,0899

Fig 9 Output simulazione Var Latenti Caso Generale

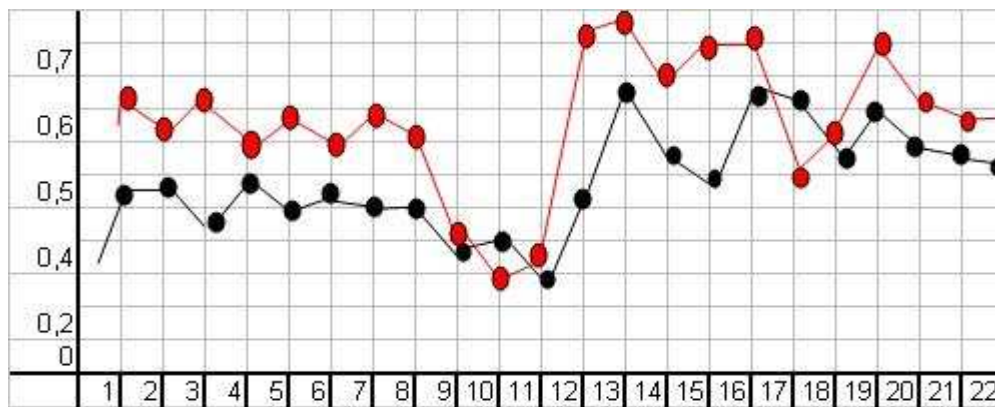


6.7.1. Confronto MSE Variabili Manifeste: CASO GENERALE

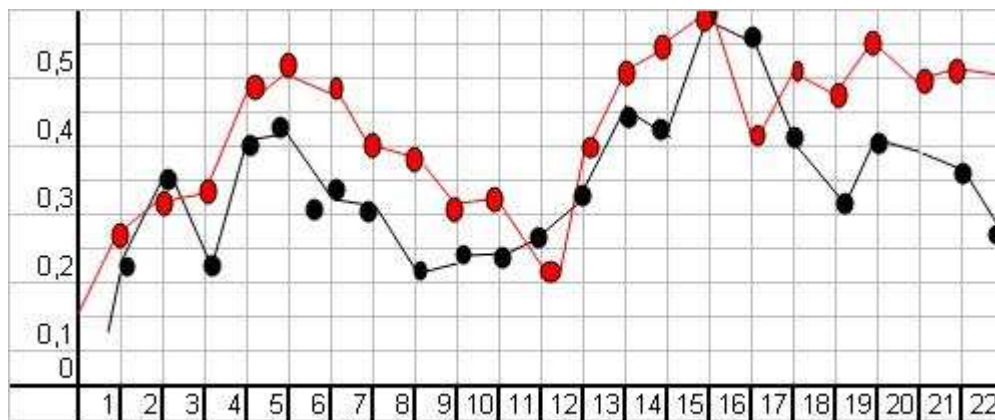
VARIABILI MANIFESTE CASO GENERALE		15		30		50	
		GME	PLS	GME	PLS	GME	PLS
P1		0,5233	0,6565	0,2144	0,299	0,0036	0,0079
P2		0,5264	0,6185	0,3547	0,3215	0,0452	0,0499
P3		0,4765	0,6596	0,2113	0,3275	0,0145	0,0625
P4		0,5344	0,5585	0,4112	0,4415	0,0299	0,0389
P5		0,4999	0,6144	0,4221	0,4994	0,0154	0,094
P6		0,5321	0,5998	0,4511	0,1254	0,0241	0,0456
P7		0,4985	0,6445	0,3245	0,3998	0,0045	0,0099
P8		0,5001	0,5998	0,2115	0,3788	0,0125	0,0235
P9		0,4236	0,3995	0,2587	0,3245	0,0365	0,088
P10		0,3995	0,3887	0,2541	0,3254	0,0421	0,0899
P11		0,3875	0,4265	0,2655	0,1999	0,0233	0,0355
P12		0,4987	0,7654	0,3125	0,3987	0,0125	0,0268
P13		0,6875	0,7854	0,4557	0,4995	0,0236	0,0366
P14		0,5698	0,6998	0,4112	0,4995	0,0455	0,0655
P15		0,5446	0,7654	0,6521	0,7885	0,0325	0,0421
P16		0,6887	0,7112	0,5265	0,4225	0,0568	0,0784
P17		0,5997	0,5112	0,4265	0,4995	0,0652	0,0778
P18		0,5664	0,5997	0,3215	0,4552	0,0236	0,0333
P19		0,6541	0,7101	0,4251	0,5112	0,0365	0,0452
P20		0,5645	0,6112	0,3995	0,4587	0,0116	0,0213
P21		0,5312	0,5998	0,3695	0,4785	0,0235	0,0325
P22		0,5112	0,5987	0,2665	0,4524	0,0344	0,0455

Fig10 Output simulazione Var Manifeste Caso Generale

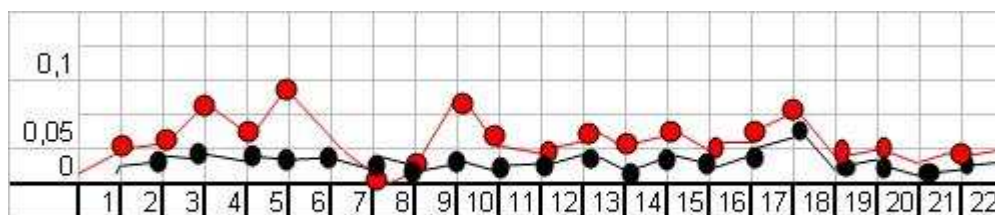
15 PLS —●— 15 GME —●—



30 PLS —●— 30 GME —●—



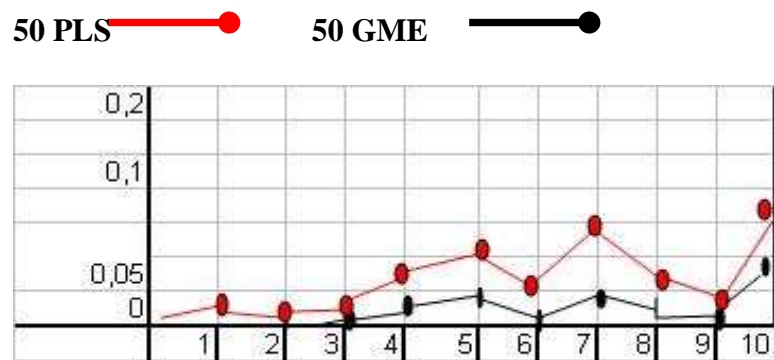
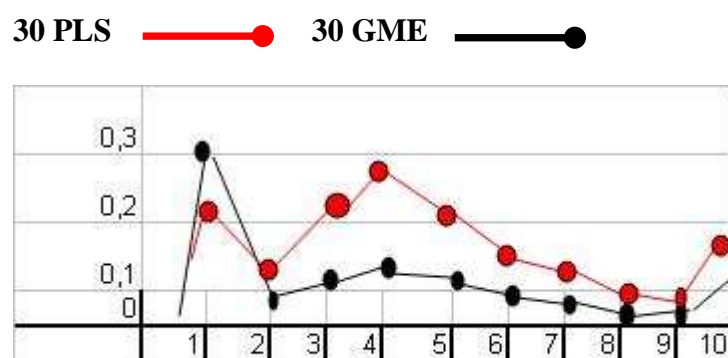
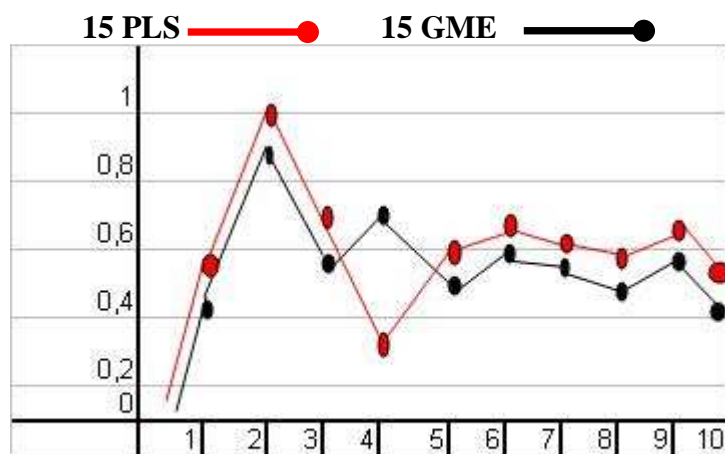
50 PLS —●— 50 GME —●—



6.8 Confronto MSE Variabili Latenti: MISSING VALUE

	15		30		50	
	GME	PLS	GME	PLS	GME	PLS
γ_1	0,3933	0,4234	0,2999	0,2164	0,0001	0,0032
γ_2	0,8597	0,9945	0,0554	0,1514	0,0009	0,0025
γ_3	0,5796	0,6532	0,1235	0,2215	0,022	0,0033
γ_4	0,6464	0,3355	0,1513	0,2855	0,0031	0,0589
γ_5	0,5758	0,5969	0,1055	0,1882	0,0054	0,064
γ_6	0,6012	0,6179	0,0421	0,0814	0,0005	0,051
γ_7	0,5898	0,5902	0,0612	0,1511	0,0069	0,0099
γ_8	0,5115	0,5995	0,0345	0,0546	0,0004	0,0594
γ_9	0,5863	0,6001	0,002	0,033	0,0017	0,0026
γ_{10}	0,4113	0,4855	0,132	0,199	0,0588	0,0899

Fig 11 Output simulazione Variabili Latenti Missing Value

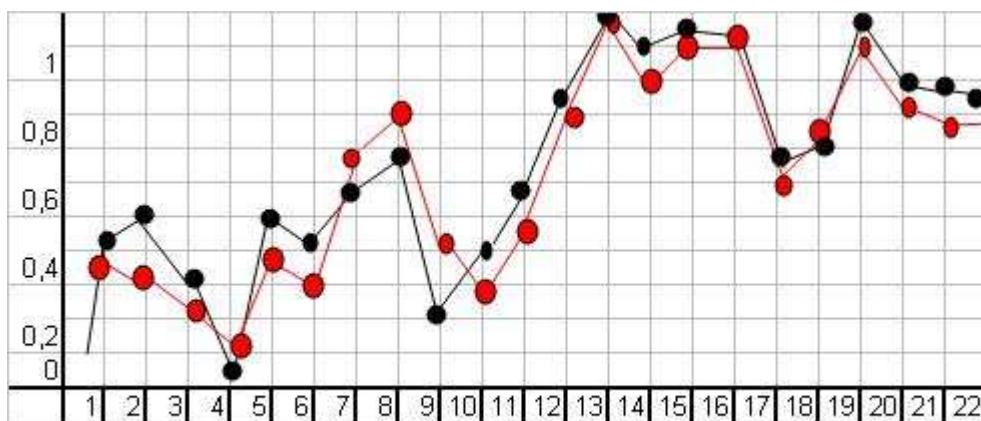


6.8.1 Confronto MSE Variabili Manifeste: MISSING VALUE

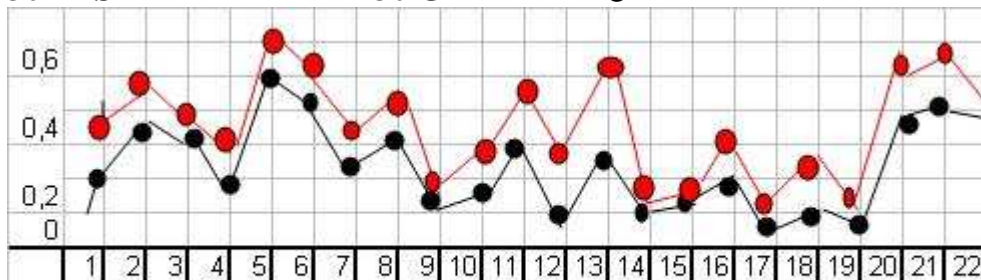
		15		30		50	
		GME	PLS	GME	PLS	GME	PLS
VARIABILI MANIFESTE	P1	0,4552	0,4123	0,3323	0,4114	0,0011	0,0055
	P2	0,5651	0,3989	0,4323	0,5685	0,0016	0,0059
	P3	0,3999	0,2676	0,3443	0,4323	0,0021	0,0053
	P4	0,1998	0,2342	0,2786	0,3998	0,0024	0,0049
	P5	0,5892	0,4453	0,5445	0,6556	0,0012	0,0041
	P6	0,4565	0,3879	0,5323	0,6112	0,0019	0,0039
	P7	0,4985	0,6112	0,3343	0,4332	0,0022	0,0043
	P8	0,6556	0,8563	0,2115	0,4998	0,0031	0,0044
	P9	0,2556	0,4673	0,2565	0,2998	0,0039	0,0046
	P10	0,3995	0,3123	0,3554	0,3998	0,0051	0,1001
	P11	0,4667	0,3223	0,3776	0,4998	0,0099	0,2001
	P12	0,6112	0,5231	0,1224	0,3987	0,1002	0,1994
	P13	0,8678	0,8556	0,3112	0,6112	0,1009	0,1999
	P14	0,9998	0,7887	0,2276	0,3132	0,1115	0,1002
	P15	1,0031	0,9991	0,2743	0,2676	0,0054	0,1005
	P16	1,0392	1,1531	0,3312	0,4112	0,0044	0,9992
	P17	0,6655	0,5898	0,1298	0,2331	0,0052	0,0766
	P18	0,9982	0,8776	0,2215	0,2887	0,0061	0,1032
	P19	0,8852	0,9121	0,1251	0,2332	0,0069	0,1023
	P20	1,0023	0,8567	0,4554	0,5998	0,1001	0,1112
	P21	0,9562	0,8773	0,5695	0,6112	0,1009	0,3012
	P22	0,8851	0,7998	0,5001	0,5112	0,1221	0,3123

Fig12 Output simulazione Var Manifeste: Missing Value

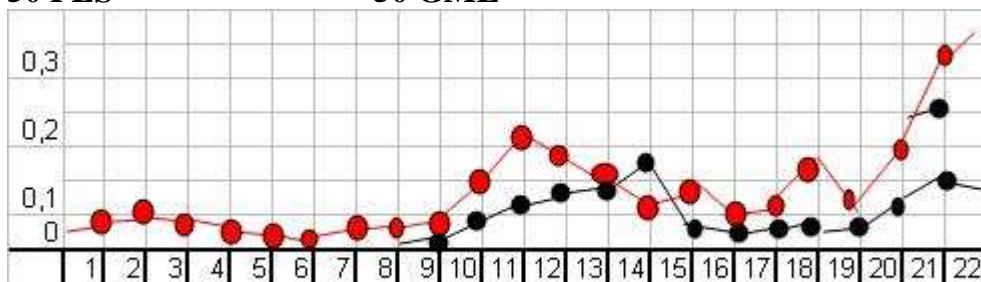
15 PLS —●— 15 GME —●—



30 PLS —●— 30 GME —●—



50 PLS —●— 50 GME —●—

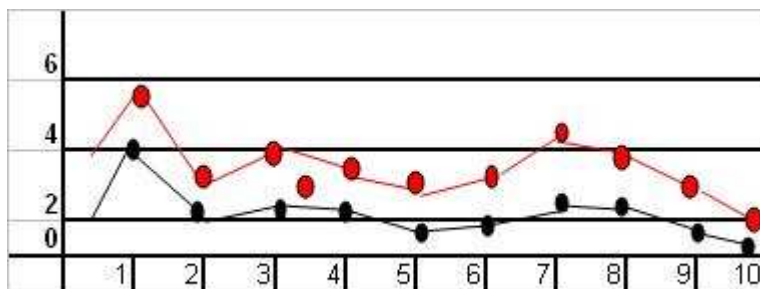


6.9 Confronto MSE Variabili Latenti: OUTLIERS

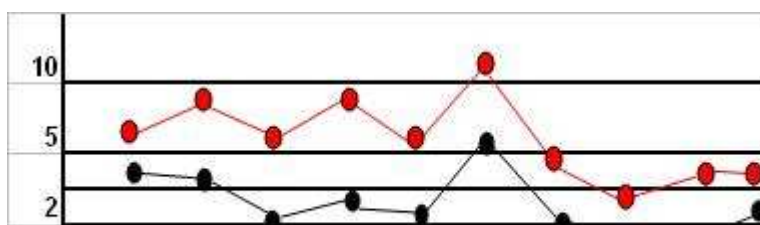
Var latenti						
	15		30		50	
	GME	PLS	GME	PLS	GME	PLS
y_1	3,995	4,864	3,223	5,332	0,2334	0,395
y_2	1,676	2,546	4,223	6,445	0,2009	0,334
y_3	2,4554	4,003	3,676	5,676	0,334	0,4322
y_4	2,113	3,335	3,454	4,998	0,1998	0,2444
y_5	0,088	2,256	3,879	5,323	0,2334	0,2445
y_6	0,998	3,334	8,565	11,009	0,1787	0,1986
y_7	2,546	4,443	2,001	6,565	0,1342	0,211
y_8	2,332	3,999	2,434	3,001	0,1023	0,1967
y_9	0,8778	2,998	1,998	4,332	0,198	0,1999
y_{10}	0,6776	0,999	2,443	3,009	0,2323	0,3223

Fig 13 Output Simulazione Variabili Latenti: Outliers

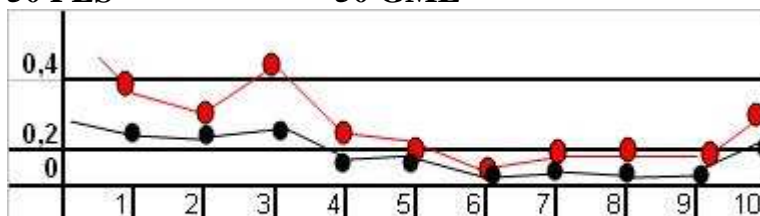
15 PLS —●— 15 GME —●—



30 PLS —●— 30 GME —●—



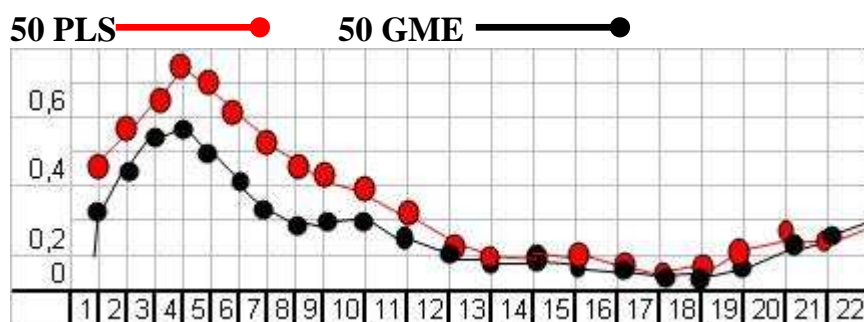
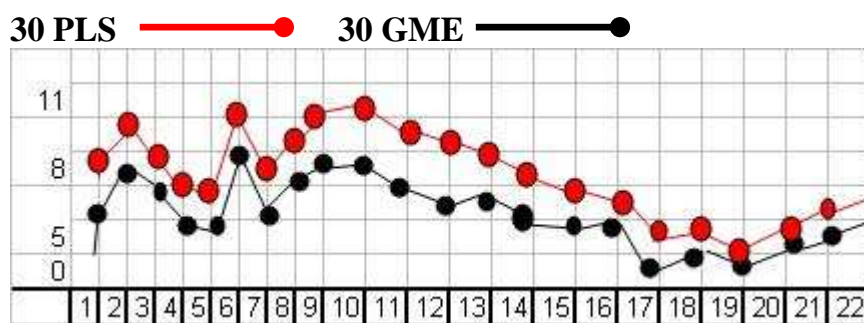
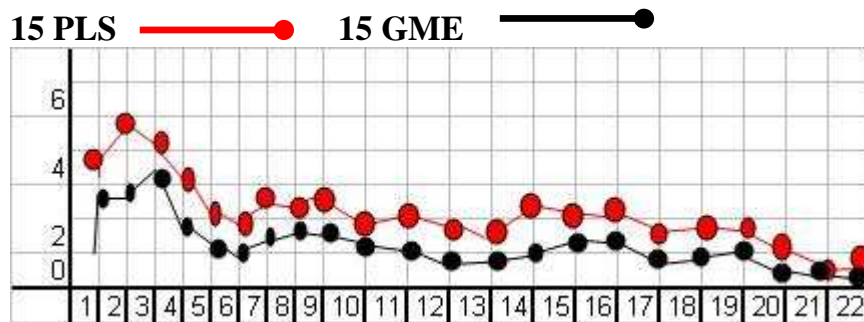
50 PLS —●— 50 GME —●—



6.9.1 Confronto MSE Variabili Manifeste: OUTLIERS

		15		30		50	
		GME	PLS	GME	PLS	GME	PLS
VARIABILI MANIFESTE OUTLIERS	P1	3,545	4,2322	6,1231	8,2321	0,2345	0,4321
	P2	3,5676	5,9781	8,3425	10,3223	0,4241	0,5998
	P3	4,0124	5,3231	7,9981	9,3231	0,4998	0,6545
	P4	2,8998	4,1221	5,9992	8,7872	0,5673	0,7881
	P5	2,2231	2,993	5,7746	6,9831	0,5231	0,6998
	P6	2,0039	2,6875	9,3312	11,3211	0,4123	0,6445
	P7	2,7873	3,1221	7,7774	8,9918	0,3892	0,6122
	P8	2,8778	3,6877	7,9983	9,4311	0,3565	0,5621
	P9	2,9889	3,8778	8,4536	10,4672	0,3443	0,4998
	P10	2,1123	3,2212	8,3425	9,1231	0,36673	0,3998
	P11	2,1009	3,3321	7,9984	10,1231	0,2435	0,3554
	P12	1,8372	3,1121	7,5632	9,9812	0,2112	0,2988
	P13	1,9565	2,9983	7,7758	9,3122	0,1342	0,1563
	P14	1,9998	3,4332	6,4352	8,8998	0,0764	0,0872
	P15	2,1123	3,2231	5,5655	7,0123	0,0556	0,0981
	P16	2,2234	3,3289	5,4353	6,8981	0,0755	0,0786
	P17	1,995	2,5664	2,9982	5,7874	0,0433	0,0456
	P18	2,0098	2,8998	4,1215	5,3321	0,0599	0,0699
	P19	2,0012	2,7878	3,9984	5,1213	0,0989	0,0982
	P20	0,8567	2,6673	4,6775	4,9981	0,1998	0,2312
	P21	0,9981	1,9983	4,9992	5,4534	0,2001	0,2112
	P22	0,7998	1,6776	5,3424	5,9982	0,2321	0,2423

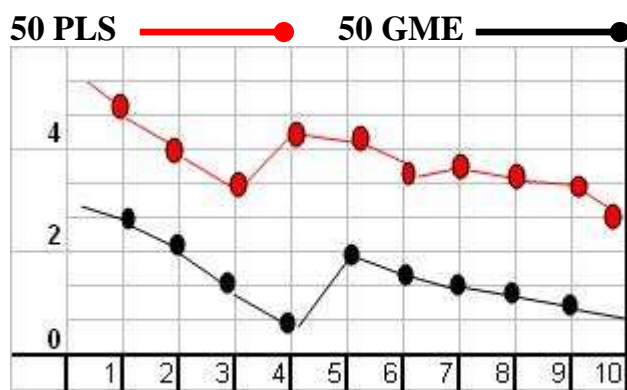
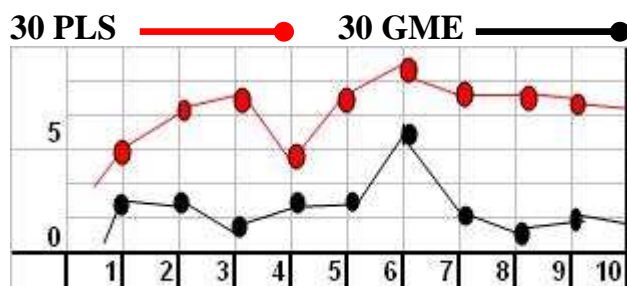
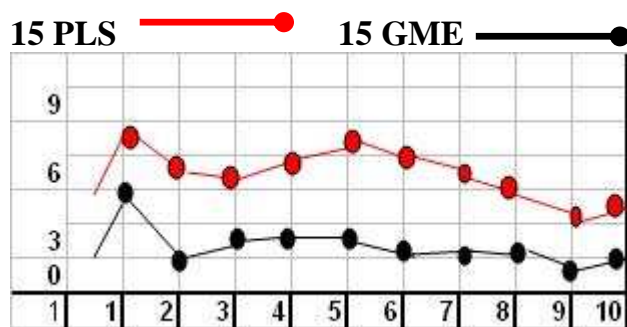
Fig 14 Output simulazione Variabili Manifeste: Outliers



6.10 Confronto MSE variabili Latenti: MULTICOLLINEARITÀ 0,3 (bassa)

	15		30		50	
	GME	PLS	GME	PLS	GME	PLS
γ_1	3,995	7,434	2,112	5,109	2,998	4,775
γ_2	1,676	6,442	3,223	5,998	2,554	3,998
γ_3	2,009	5,998	1,898	7,564	0,998	3,554
γ_4	2,113	7,399	3,223	3,998	0,556	4,554
γ_5	1,788	8,456	3,879	8,998	2,667	4,223
γ_6	0,998	7,334	5,778	7,565	1,998	3,598
γ_7	2,4332	5,443	3,562	6,998	1,887	3,998
γ_8	1,998	4,954	2,113	6,7678	1,566	3,565
γ_9	1,655	5,004	3,998	5,998	1,434	3,232
γ_{10}	1,987	5,3998	2,443	5,556	1,009	1,889

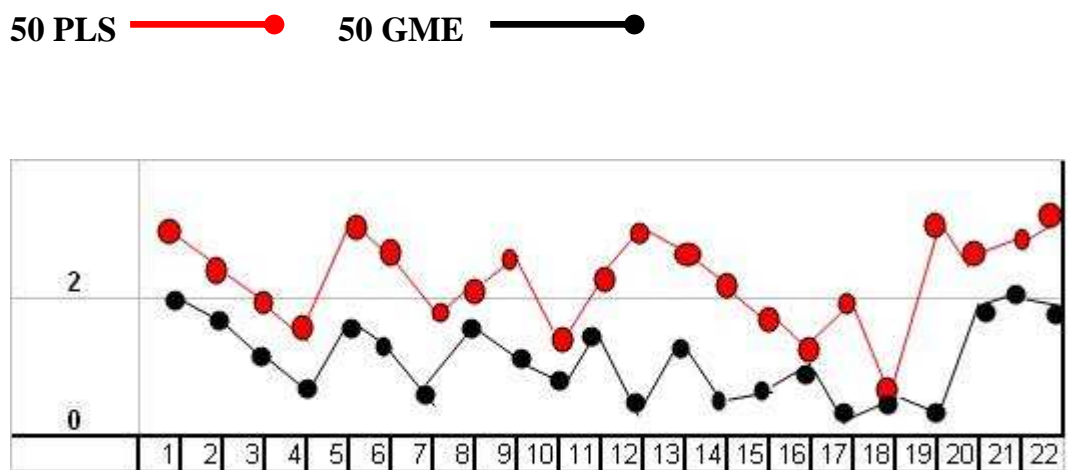
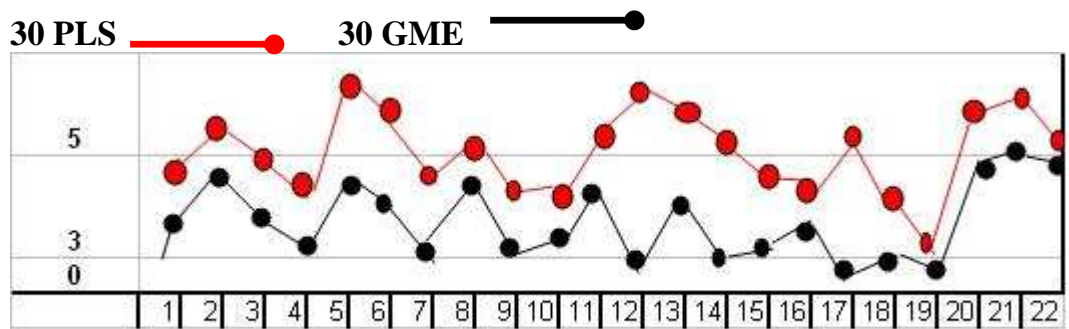
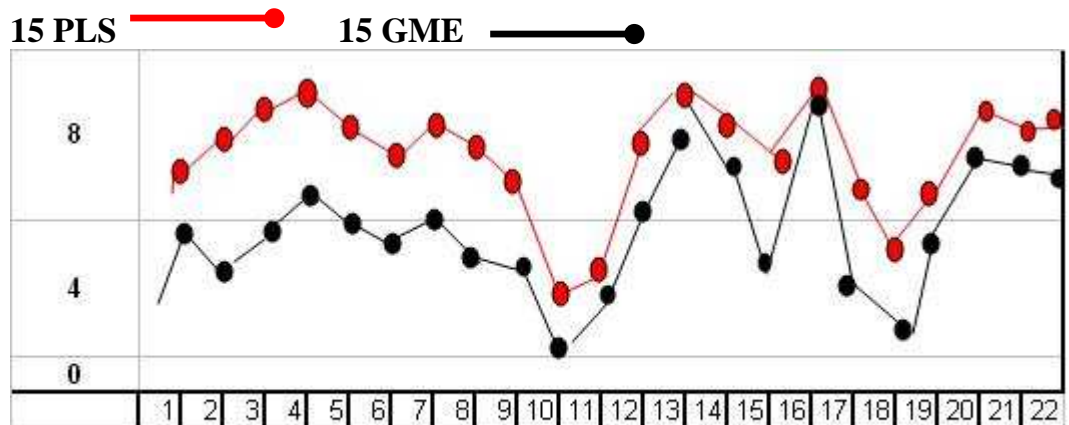
Fig 15 Output simulazione Variabili Latenti:
Multicollinearità bassa



**6.10.1 Confronto MSE Variabili Manifeste:
MULTICOLLINEARITÀ 0,3 (bassa)**

		15		30		50	
		GME	PLS	GME	PLS	GME	PLS
VARIABILI MANIFESTE MULTICOLLINEARITÀ	P1	6,3255	8,3255	2,2385	4,2554	2,0121	3,9541
	P2	5,2132	8,8954	4,5452	6,2651	1,9985	3,6521
	P3	5,9961	9,0214	2,5564	5,1235	1,7575	2,2352
	P4	6,9981	9,5457	1,998	4,5451	1,3257	1,9563
	P5	5,3251	8,5875	4,9886	6,9868	1,8756	4,0212
	P6	4,998	7,9985	4,9111	6,5654	1,9654	3,996
	P7	6,0101	8,9656	1,8758	5,0112	1,2321	2,1123
	P8	5,0121	8,0232	5,001	7,0122	1,8778	2,332
	P9	4,8654	7,6542	1,9986	4,9987	1,9554	4,001
	P10	4,2123	5,0566	2,1011	5,0123	1,8998	1,9975
	P11	4,5421	4,9986	5,1231	6,5111	1,8998	3,8775
	P12	5,998	9,4532	1,7877	7,1122	1,1235	4,1232
	P13	8,8875	9,8786	5,0001	6,3253	1,9002	4,022
	P14	9,6545	9,7996	1,9154	6,4541	1,9965	3,8542
	P15	4,9986	9,1021	2,0001	5,2123	1,1123	1,5654
	P16	9,7565	9,9986	2,6695	5,1325	1,4521	1,4565
	P17	4,4545	7,6332	1,6587	6,6571	0,9874	1,9665
	P18	4,5655	5,0121	1,9653	4,998	1,0232	1,2321
	P19	5,1235	7,7652	1,8756	2,2335	0,9887	4,2214
	P20	8,8985	9,1235	4,6521	6,2112	2,1201	4,0121
	P21	8,5465	8,2321	4,8788	6,3244	2,3120	4,1955
	P22	7,9886	8,2321	4,6521	6,6554	2,2012	4,2001

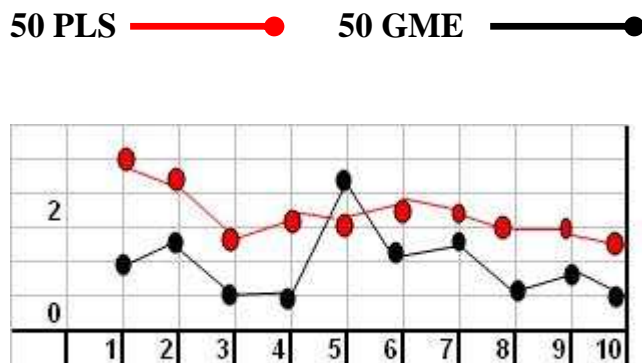
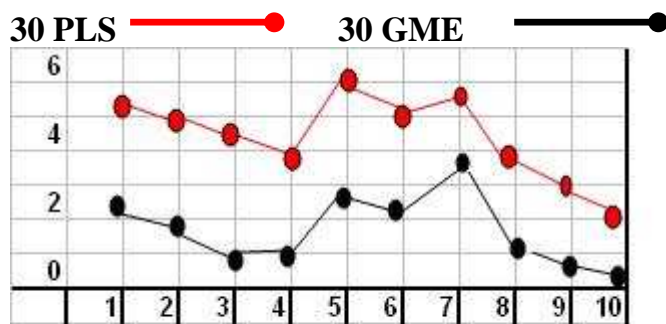
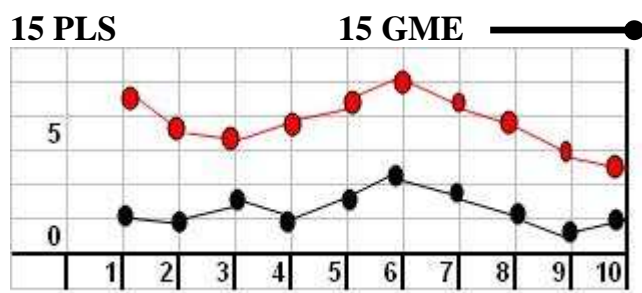
Fig 16 Output simulazione Variabili Manifeste: Multicollinearità bassa



6.11 Confronto MSE Variabili Latenti: MULTICOLLINEARITÀ 0,9 (alta)

	15		30		50	
	GME	PLS	GME	PLS	GME	PLS
γ_1	1,2554	6,5235	2,1912	4,4458	1,2161	2,97
γ_2	0,9865	5,4565	2,1256	4,0986	1,6144	2,4751
γ_3	1,5665	5,0121	1,0125	3,5632	0,6851	1,2830
γ_4	1,1225	5,3652	1,0236	3,0875	0,6862	1,6222
γ_5	1,2102	5,5562	2,5322	5,5002	2,6009	1,4625
γ_6	1,6852	6,6891	1,7254	4,0818	1,0602	1,9932
γ_7	1,5652	5,6985	3,65254	4,9669	1,2116	1,8767
γ_8	1,2236	5,5696	1,0625	2,1984	0,3214	1,4952
γ_9	0,9986	5,2362	0,9687	1,9624	0,4254	1,3254
γ_{10}	1,1256	3,3654	0,6587	1,8685	0,3112	1,1002

Fig 17 Output simulazione Variabili Latenti: Multicollinearità 0,9 (alta)

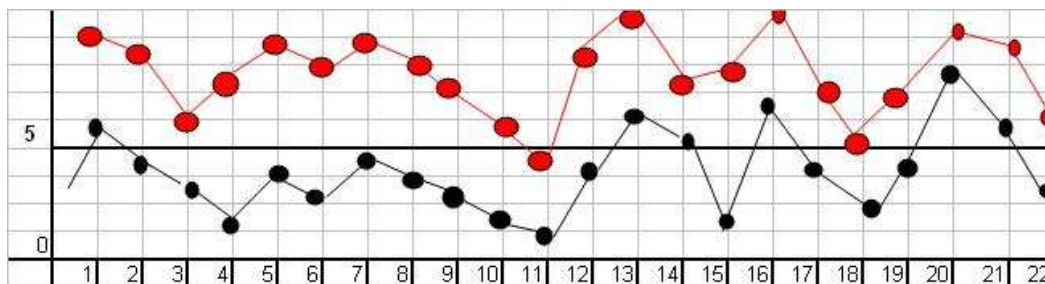


6.11.1 Confronto MSE Variabili Manifeste: MULTICOLLINEARITÀ 0,9 (alta)

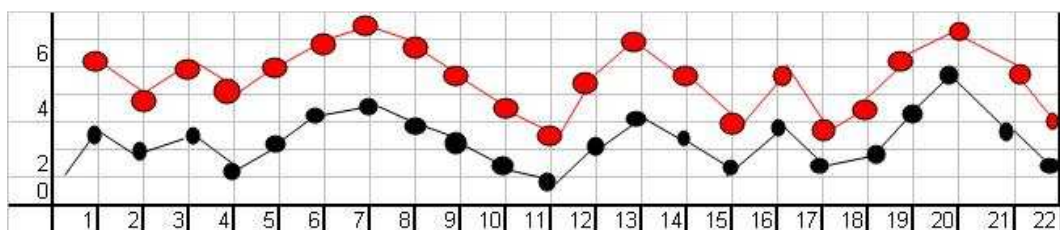
		15		30		50	
		GME	PLS	GME	PLS	GME	PLS
VARIABILI MANIFESTE MULTICOLLINEARITÀ' 0,9	P1	5,6492	7,3471	3,5652	6,265	1,699	2,2365
	P2	4,325	6,2541	3,1214	5,2365	2,1225	2,5665
	P3	3,2541	5,2654	3,3215	6,1225	1,2354	2,0223
	P4	2,1263	5,9654	2,0121	5,5565	0,7885	1,9987
	P5	3,5452	6,1987	3,4554	5,9985	1,8756	2,6653
	P6	3,7485	6,0121	4,1211	6,5654	1,5332	2,2256
	P7	3,8987	6,2112	4,2321	7,0121	0,9986	1,9986
	P8	4,4565	6,0022	3,8985	6,2322	1,5365	2,5563
	P9	3,8762	7,6542	3,5654	5,9985	0,7885	1,8876
	P10	3,2154	5,0566	2,2145	4,8988	1,2136	2,5563
	P11	2,0212	4,9986	1,9985	3,7887	0,6885	1,5665
	P12	4,5654	6,1235	3,4521	5,2351	1,4235	2,4568
	P13	5,5442	5,1211	3,7887	6,5565	1,5565	3,2554
	P14	4,9985	5,3225	3,5652	5,9981	1,3265	2,6535
	P15	2,2565	5,5547	2,0001	4,1125	1,0215	2,2115
	P16	6,5565	7,9865	4,2214	5,9968	0,9886	1,6885
	P17	4,6525	5,2321	2,3554	3,9986	0,6521	1,5654
	P18	3,6574	4,5652	3,6552	3,2211	1,4565	2,1554
	P19	4,5998	4,6855	4,3652	6,2351	1,1235	2,8998
	P20	4,7854	7,1256	5,9986	6,3254	1,6523	2,9975
	P21	6,5411	6,1254	4,0235	5,9986	1,4996	2,8756
	P22	3,8754	4,9652	2,2236	3,9952	0,8775	2,0121

Fig 18 Output simulazione Variabili manifeste: Multicollinearità 0,9 (alta)

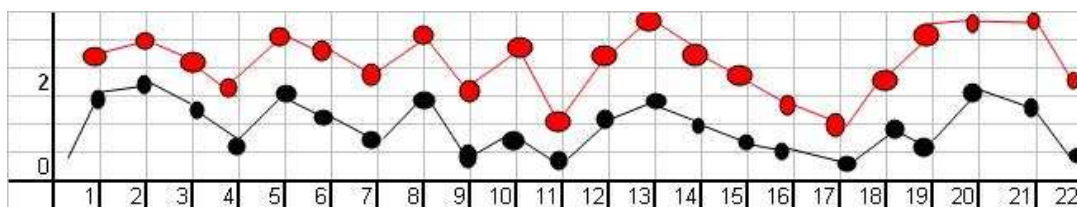
15 PLS —●— 15 GME —●—



30 PLS —●— 30 GME —●—



50 PLS —●— 50 GME —●—



6.12 Analisi dei risultati

Caso generale: i risultati ottenuti, basati sulla simulazione, considerando sia l'output delle variabili manifeste, che latenti, evidenziano una migliore performance della metodologia GME rispetto al PLS nella stima dei parametri del modello.

Missing Value: Sia con i risultati inerenti le variabili manifeste, che in quelli con le variabili latenti, la performance GME risulta essere migliore rispetto a PLS, tuttavia per campioni di piccole dimensioni ($N=15$) si evidenzia un lieve miglioramento nella stima ottenuta con la metodologia PLS.

Outliers: La performance secondo la metodologia GME per campioni di dimensione $N=15, 30$ (sia nel caso di variabili latenti che manifeste) è preferita rispetto a quella PLS nella stima dei parametri. Tuttavia entrambe le metodologie risultano essere asintoticamente equivalenti.

Multicollinearità: anche se la metodologia GME è caratterizzata dalla presenza di multicollinearità, i risultati dimostrano che anche in questo caso la stima GME è migliore di quella ottenuta con quella PLS.

Dallo studio effettuato con il nostro database, si evidenzia una conferma di ciò che prevale in letteratura, in effetti abbiamo una conferma che la stima GME, anche nelle nostre condizioni sperimentali, ha migliori performance migliori, anche nel caso di dati mancanti, Outliers e multicollinearità, rispetto alla stima PLS.

Il risultati che provengono tramite stima GME produce stime consistenti (nessun vincolo sui parametri).

Questo metodo è molto utile quando si hanno dati incompleti e campioni di piccole dimensioni.

APPENDICE

Bootstrap

Il Bootstrap è stato proposto da Bradley Efron nel 1979 come evoluzione del metodo jackknife, per stimare l'errore standard di un parametro della popolazione. In pochi anni, questa procedura ha avuto una evoluzione rapida e una serie di approfondimenti da parte dello stesso autore e dei suoi colleghi, che l'hanno resa la tecnica di ricampionamento più nota e diffusa.

Il nome bootstrap (letteralmente stringhe o lacci da scarpe) è derivato dall'espressione inglese "to pull oneself up by one's bootstrap" (tirarsi su attaccandosi ai lacci delle proprie scarpe), tratto dal romanzo del diciottesimo secolo "Adventures of Baron Munchausen di Rudolph Erich Raspe. Evidenzia, in modo scherzoso, il fatto paradossale che l'unico campione disponibile serve per generarne molti altri e per costruire la distribuzione teorica di riferimento.

Per il grande impegno scientifico e divulgativo dei suoi proponenti, il metodo bootstrap ha il vantaggio di fornire una serie ampia di esemplificazioni, in articoli pubblicati su riviste a diffusione internazionale, in merito all'inferenza anche per funzioni molto complesse. Oltre agli esempi indicati nella presentazione generale di questi metodi, sono casi ulteriori d'applicazione il calcolo dei momenti, dei coefficienti di variazione, dei rapporti tra valori medi e fra varianze, dei coefficienti di correlazione, degli autovalori delle matrici di varianze e covarianze.

L'uso del bootstrap non è possibile con i quantili (quanto si disponga solo di essi non dei valori reali), con dati incompleti, non indipendenti o alterati da errori grossolani.

Come il jackknife, questa tecnica permette di ricavare gli errori standard e i limiti di confidenza di varie misure statistiche, che hanno distribuzioni non note o molto complesse. E' un metodo generale per ottenere informazioni circa la variabilità e la distribuzione di statistiche campionarie $\hat{\theta}$; quindi permette di stimare i limiti di confidenza del parametro θ della popolazione, quando non si possiedono informazioni sulla sua distribuzione.

Se il campione è formato da k dati, l'idea di base è di estrarre da esso per campionamento semplice con ripetizione molti campioni di k osservazioni, allo scopo di trovare la probabilità che la misura in oggetto cada all'interno di intervalli predeterminati.

Il campione bootstrap è nient'altro che il campione originario nel quale, per effetto dell'estrazione con ripetizione, alcuni dati sono ripetuti ed altri, per mantenere lo stesso numero d'osservazioni, sono assenti. E' proprio la modalità di estrazione, fondata sulla ripetizione, a generare la variabilità nelle stime; poiché è richiesto che i campioni abbiano tutti lo stesso numero d'osservazioni, se si estraesse senza ripetizione sarebbero tutti identici.

Ognuna di queste stringhe di k osservazioni può contenere due o più valori identici, con l'ovvia esclusione d'altri valori che sono contenuti nel campione originale. Sono chiamati campioni di bootstrap, ognuno dei quali permette di ottenere una stima della statistica desiderata.

La distribuzione della misura statistica calcolata è trattata come una distribuzione costruita a partire da dati reali (cioè della popolazione) e fornisce una stima dell'accuratezza statistica.

Bibliografia:

Al-Nasser A. D. (2003). Customer Satisfaction Measurement Models: Generalized Maximum Entropy Approach. *Pak Journal of Statistics*, 19(2), 213–226.

Altieri L., *Servizio qualità efficienza sono in antitesi?*. “Logistica Management”, n. 90, novembre 1998, pp. 65-67

Amenta P. (1993). Il coefficiente di correlazione lineare tra matrici dei dati nel contesto multivariate, AMASES, Ischia, settembre 1993.

Anderson T.W., *An introduction to multivariate statistical analysis*. New York, Wiley, 1958.

Arato G., *Come conquistare il cliente. Il Modello di Kano e la creazione della “Attractive Quality”*. “Qualità”, anno XXVIII, n. 5, settembre/ottobre 1998, pp. 17-19;

Bertossi G. et al., *Ruolo del personale e Customer Satisfaction*. Cit., pp. 99-109.

- Carmines F., MCiver J., *Analyzing models with unobserved variables: analysis of covariance structures*. "Social Measurement", Beverly Hills, Sage.
- Ciavolino E. (2007). The Entropy Theory for evaluating the Job Satisfaction, *GFKL 2007*, Freiburg, march 2007.
- Cappelli L., Riccio A.V., *Customer satisfaction e imprenditore schumpeteriano: conflitti e convergenze*. In *Qualità: cultura per il 2000 - Competizione globale e progresso civile*. XX Convegno nazionale dell'AICQ, atti del convegno, vol. C, Bologna, 15-17 maggio 2000, pp. 250-264.
- Chiacchierini E., *Tecnologia e produzione*. ed. Kappa, Roma, 1996.
- Ciavolino E., D'Ambra L., De Franco D., (2007). *Massima Entropia Generalizzata e Simple Component Analysis per l'analisi della Customer Satisfaction nel Trasporto Pubblico Locale: L'Esperienza dell'Azienda AMTS di Benevento*. Rivista Di Economia E Statistica Del Territorio. vol. 3, pp. 201-221, ISSN: 1971-0380.
- Ciavolino E., Dahlgaard J.J., (2009), *Simultaneous Equation Model based on Generalized Maximum Entropy for studying the effect of the Management's Factors on the Enterprise Performances*, Journal Of Applied Statistics, Vol. 36, Issue 7, pp. 801 - 815, ISSN: 1360-0532.
- Ciavolino E., Al Nasser A.D., D'Ambra A. (2006). The Generalized Maximum Entropy Estimation method for the Structural Equation Models, *GFKL 2006*, Berlino, marzo.
- Ciavolino Et., AL-Nasser A.D., (2009), *Comparing Generalized Maximum Entropy and Partial Least Squares methods for Structural Equation Models*, In press on Journal of Nonparametric Statistics
- Coen G. et al, *Customer Satisfaction nell'azienda di trasporto pubblico locale di Campobasso*. "Qualità", anno XXIX, n. 8, settembre 1999, pp. 66-79.
- Costabile M., *La misurazione della Customer Satisfaction: nuove ipotesi sul paradigma conferma/disconferma*. "Micro & macro marketing", anno V, n. 3, dicembre 1996, pp. 475-501.
- Corbetta P., *Metodi di analisi multivariata per le scienze sociali*. Cit., p. 122
- Cronin, J. J., & Taylor, S. A., *SERVPERF versus SERVQUAL: reconciling performance-based and perceptions-minus-expectations measurement of service quality*. "Journal of Marketing", vol. 58, gennaio 1994.

- Cronin, J. J., & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension, *Journal of Marketing*, 56 (July). 55-68.
- D'Ambra L. *Lezioni di inferenza statistica*. Napoli, Rocco Curto Editore, 2000;
- D'Ambra L. & Gallo M. (2006). La valutazione Customer Satisfaction: aspetti statistici e recenti sviluppi, *Metodi di Valutazione ex post dei servizi di Pubblica Utilità alla persona: Uno sguardo di Insieme*; edizioni Guerini.
- D'Ambra L., Ciavolino E. (2007). Riflessioni per lo sviluppo di un sistema integrato di Customer Satisfaction per il Trasporto Pubblico Locale, in Pubblicazione *DASES Franco Angeli 2007*, in stampa.
- Escoufier Y. (1973). Le traitement des variables vectorielles, *Biometrics*, 1973.
- Farnum N. R., *Modern statistical quality control and improvement*. Belmont, Duxbury Press, 1997.
- Franceschini, F., Rossetto, S. (1997). On-line service quality control: the 'Qualitometro' method, *De Qualitate*, Vol. 6 No.1, pp.43-57.
- Franceschini, F. Stangalini M., *Un'applicazione del metodo "Qualitometro" per la valutazione della qualità*. "De Qualitate", anno IX, n. 4, aprile 2000, pp. 55-65;
- Franceschini, F., Rossetto, S. *Qualità nei servizi: un metodo per la valutazione e il controllo in linea del differenziale tra Qualità attesa e percepita*. "De Qualitate", anno V, n. 3, marzo 1996, pp. 53-64.
- Franceschini, F., Rossetto, S., *La valutazione e il controllo in linea della qualità dei servizi*. "De Qualitate", anno VI, n. 1, gennaio 1997, pp. 43-57.
- Franceschini, F., Rossetto, S. (1997). On-line service quality control: the 'Qualitometro' method, *De Qualitate*, Vol. 6 No.1, pp.43-57.
- Golan A., Judge G. & Karp L. (1996). A maximum entropy approach to estimation and inference in dynamic models or counting fish in the sea using maximum entropy, *Journal of Economic Dynamics and Control*, 20, 559-582.
- Goldberger A.S., Duncan O.D., *Structural Equation Models in the Social Science*. New York, Seminar Press, 1973
- Gualtieri F., Potiè C., *Dall'assicurazione della qualità alla fidelizzazione del cliente*. "De Qualitate", anno X, n. 5, maggio 2000, pp. 19-31
- Hinterhuber H. et Al, *Un modello semiquantitativo per la valutazione della soddisfazione del cliente*. "Micro & macro marketing", anno VI, n. 1, aprile 1997, pp. 127-143

- Jaynes E.T. (1957). Information Theory and Statistical Mechanics, *The Physical Review* 106 (4), 620-630, May 15, 1957.
- Jaynes, E. T. (1968). Prior Probabilities, *IEEE Transactions On Systems Science and Cybernetics*, vol. sec-4, no. 3, 227-241.
- Joliffe I.T (2002). Principal Component Analysis (2nd ed.). Springer, New York.
- Jöreskog K.G. (1970). A general method for estimating a Linear Structural Equation System, in *Goldberger e Duncan*, 85-112.
- Kano N., *Business strategies for the 21st century and attractive quality creation, proceeding of international conference for quality*. Yokohama, 1996, pp. 105-108
- Londrillo L., *Il sistema Servqual come strumento di valutazione sulla qualità interna all'azienda*. "De Qualitate", anno VIII, n. 10, novembre 1999, pp. 39-49
- Oliveti A., *La valutazione della qualità dei servizi sanitari: dalla descrizione alla spiegazione*. In *Atti del I Convegno Internazionale: Gestione della Qualità Totale nelle strutture sanitarie: dalla teoria alla pratica*. Troina (EN), 17 aprile 1998, pp. 1-14.
- Orsingher C., *Strumenti di misurazione della Customer Satisfaction e della qualità nelle imprese di servizi: una rilettura critica*. "Sinergie", anno XIV, n. 40, maggio/agosto 1996, pp. 209-225.
- Parasuraman A., *Servqual: applicazioni, messa a punto sul campo e risultati*. "De Qualitate", anno V, n. 6, giugno 1996, pp. 5-7.
- Parasuraman A., Berry L.L., Zeithaml V.A. (1991). Refinement and Reassessment of the Servqual Scale, *Journal of Retailing*, Vol. 67, 420-450.
- Parasuraman A., Zeithaml V.A., Berry L.L. (1988). Servqual: A Multiple-Item Scale for Measuring Customer Perceptions of Service Quality, *Journal of Retailing*, Vol. 64, 11-40.
- Parasuraman A.m Zeithalm V.A., Berry L.L., *A conceptual model of service quality and its implications for future research*. "Journal of Marketing", vol. 49, 1985, pp. 41-50.
- Persico P., Vinci S., *Principi di Econometria*. Napoli, Liguori, 1981.
- Popper K. R., *The logic of scientific discovery*. New York, Basic Books, 1959; trad. it., *Logica della scoperta scientifica*. Torino, Einaudi, 1970

- Rizzi A., *Il linguaggio delle matrici. Le applicazioni in economia, in statistica e nelle scienze sociali*. Roma, La Nuova Italia Scientifica, 1988;
- Schvaneveldt S. J., Enkawa T., Miyikawa M., *Consumer evaluation perspectives of service quality: evaluation factors and two-way model of quality*. "Total Quality Management", vol. 2, n. 2, 1991
- Scarpinato M., *La valutazione della Customer Satisfaction come strumento di analisi strategica*. "Sviluppo & Organizzazione", n. 141, gennaio/febbraio 1994, pp. 2-6
- Schumpeter J.A., *The theory of Economic Development*. Harvard University Press, Cambridge, 1934.
- Stevens S.S., *On the theory of scales of measurement*. "Science", n. 103, 1946, pp. 670-680
- Shannon C. E. (1948). A mathematical Theory of Communications, *Bell System Technical Journal*, 27, 379–423.
- Tassinari G. Et al., *La soddisfazione del cliente dei servizi di segreteria universitaria: un modello ad equazioni strutturali*. In *Valutazione della qualità e Customer Satisfaction: il ruolo della statistica. - Aspetti oggettivi e soggettivi della Qualità*. Atti della Giornata di studio promossa dall'AICQ, Bologna 24 settembre 1999, pp. 291-316.
- Teas R. K., *Expectations, performance, evaluation and consumers perceptions of Quality*. "Journal of Marketing", vol. 57, luglio 1993
- Tenenhaus M., Esposito Vinzi V., Chatelin Y.M. and Lauro C. (2005) *PLS path modeling Computational Statistics & Data Analysis*, Volume 48, Issue 1, pp. 159-205
- Tronconi A., *Il controllo di business nell'impresa di servizi*. "De Qualitate", anno V, n. 6, giugno 1996, pp. 55-64
- Valdani E., Busacca B., *La Customer Satisfaction: specificità, analisi e management*. "Micro & macro marketing", anno IV, n. 3, dicembre 1995, pp. 315-343
- Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Jöreskog and H. Wold (Eds.), *Systems under Indirect Observations II* (pp. 1–54). Amsterdam: North-Holland.
- Zanella A., Cerri M., *La misura di Customer Satisfaction: qualche riflessione sulla scelta delle scale di punteggio*. In *Valutazione della qualità e Customer*

Satisfaction: il ruolo della statistica. - Aspetti oggettivi e soggettivi della Qualità.
Atti della Giornata di studio promossa dall'AICQ, Bologna, 24 settembre 1999,
pp. 217-231.